AD_____

Award Number: W81XWH-06-1-0281

TITLE: Simultaneous Analysis of Germline Cnps and Snps in High Risk Prostate Cancer Families

PRINCIPAL INVESTIGATOR: Jianfeng Xu, M.D., Ph.D.

CONTRACTING ORGANIZATION: Wake Forest University Health Sciences
Winston-Salem, NC 27157

REPORT DATE: October 2007

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

DISTRIBUTION STATEMENT: Approved for Public Release;
Distribution Unlimited

The views, opinions and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy or decision unless so designated by other documentation.

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching existing data sources, gathering and maintaining the data needed, and completing and reviewing this collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Department of Defense, Washington Headquarters Services, Directorate for Information Operations and Reports (0704-0188), 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302. Respondents should be aware that notwithstanding any other provision of law, no person shall be subject to any penalty for failing to comply with a collection of information if it does not display a currently valid OMB control number. **PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.**

1. REPORT DATE *(DD-MM-YYYY)*	**2. REPORT TYPE**	**3. DATES COVERED** *(From - To)*
01-10-2007	Final	10 Jan 2006 - 30 Sep 2007

4. TITLE AND SUBTITLE

Simultaneous Analysis of Germline Cnps and Snps in High Risk Prostate Cancer Families

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-06-1-0281

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)
Jianfeng Xu, M.D., Ph.D.

E-Mail: jxu@wfubmc.edu

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Wake Forest University Health Sciences
Winston-Salem, NC 27157

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

Prostate cancer is the leading cancer among men in the United States, and is a disease with strong genetic susceptibility. The genetic susceptibility is due to the inheritance of altered germline DNA sequences, either in the form of point mutations such as single nucleotide polymorphisms (SNPs), or deletions/gains of a string of nucleotides such as copy number polymorphisms (CNPs). Most current genetic studies focus only on the role of SNPs in genetic susceptibility. In contrast, few studies have explored the role of deletions/gains in cancer predisposition, due to limited methods. In fact, germline deletions/gains are common in the human genome and may have a significant impact on gene products because they can involve an entire gene or a significant portion of a gene. They may play a more important role in hereditary PCa, a type of PCa that is likely due to germline changes in major genes. With the grant support, we have made important progresses toward this new research area. To our knowledge, our study is the first of its kind. Results will likely contribute to our understanding of prostate cancer etiology, and provide novel targets for prostate cancer risk assessment, prevention, and therapy.

15. SUBJECT TERMS
Prostate cancer, copy number polymorphisms (CNPs), SNPs, hereditary

16. SECURITY CLASSIFICATION OF:			**17. LIMITATION OF ABSTRACT**	**18. NUMBER OF PAGES**	**19a. NAME OF RESPONSIBLE PERSON**
a. REPORT	**b. ABSTRACT**	**c. THIS PAGE**			USAMRMC
U	U	U	UU	65	**19b. TELEPHONE NUMBER** *(include area code)*

Table of Contents

Introduction

Prostate cancer (PCa) is the leading cancer among men in the United States, and is a disease with strong genetic susceptibility. The genetic susceptibility is due to the inheritance of altered germline DNA sequences, either in the form of point mutations such as single nucleotide polymorphisms (SNPs), or deletions/gains of a string of nucleotides such as copy number polymorphisms (CNPs). Most current genetic studies focus only on the role of SNPs in genetic susceptibility. In contrast, few studies have explored the role of deletions/gains in cancer predisposition, due to limited methods. In fact, germline deletions/gains are common in the human genome and may have a significant impact on gene products because they can involve an entire gene or a significant portion of a gene. They may play a more important role in hereditary PCa (HPC), a type of PCa that is likely due to germline changes in major genes.

With the support of an Exploration-hypothesis development (EHD) grant from the DOD, we have made important progresses toward this new research area.

Body

The novel hypothesis of the proposal is that germline gross deletions/insertions (CNPs), as well as single nucleotide substitutions (SNPs), in the genome affect the function and/or expression of PCa related genes and thus contribute to the genetic susceptibility of PCa.

We had two specific aims: 1) screen for germline CNPs in the genome by measuring the allele intensity of 500K SNPs; 2) to compare the germline CNPs among relatives; 3) confirm the CNPs using real-time PCR; and 4) test for co-segregation of the identified CNPs with prostate cancer in these three families.

We have completed all the specific aims. Some of the results are published in our recent paper (Liu 2006). They are briefly described below.

Using the Affymetrix 100K SNP Mapping set to detect copy number differences. We genotyped DNA samples isolated from blood samples of 23 subjects in four PCa families ascertained at Johns Hopkins Hospital. The average genotype call rate of the 100K panel in these subjects was 99.2%. These data suggest that DNA samples isolated from blood many years ago are stable for Affymetrix SNP analysis.

We began with lymphoblastoid cell (LCL) line DNA from a PCa patient (005-015). Allele intensities of more than 100K SNPs were analyzed using default settings of CNAT (500-kb smooth averages) and CNAG. Two large-scale deletions and four large-scale gains were evident in chromosome 9 (Fig 1). For example, at the 9q31.1 region, 65 consecutive SNPs spanning 2.4 Mb signaled a deletion: copy number (CN) ≤ 1.2 and P-values ≤ $10^{-5.5}$ (Fig 2a). Similarly, at 9q21.2, a set of 24 consecutive SNPs spanning 0.5 Mb revealed a gain: CN ≥ 2.8 and P-values ≤ 10^{-4} (Fig 2b). The deletion and gain were confirmed by qPCR (Fig 2c-d). These data suggest that the Affymetrix 100K SNP panel is able to detect large-scale deletions and insertions. Furthermore, these data provide an empirical basis for establishing the criteria that can be used to define deletions and gains.

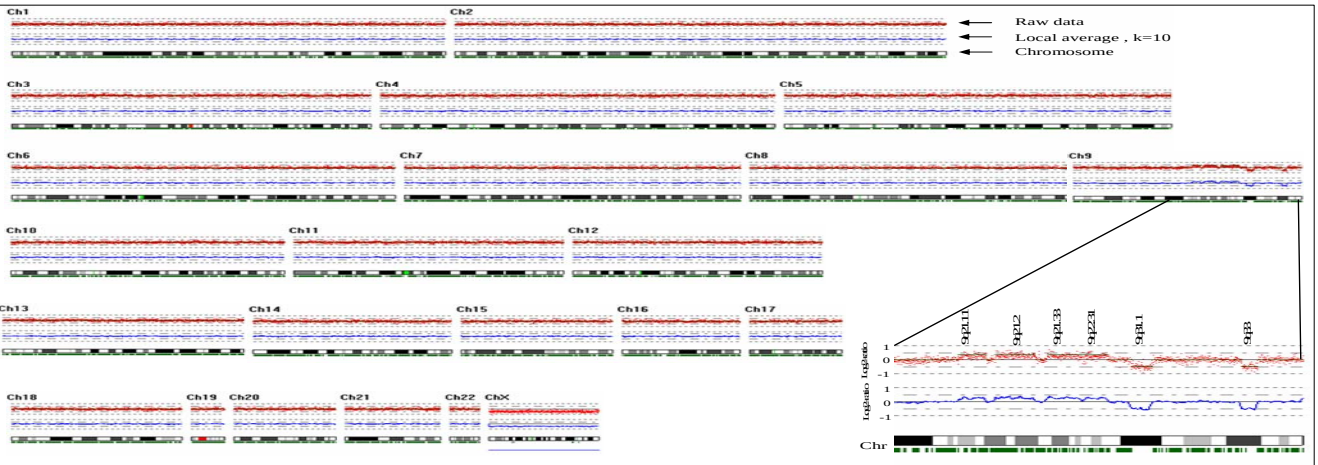

Fig 1. Genome-wide copy number analysis of GeneChip 50K Xba array data of 005-015 lymphoblastoid cell line DNA using CNAG. Chr, chromosome. Red dots, raw log2 ratio for each SNP. Green line, inferred copy number from a hidden Markov model. Blue curves, local mean analysis of ten consecutive SNPs. Green bar, heterozygous SNPs. Green bar, heterozygous SNP calls. LOH is marked under chromosome with blue lines.

Potential artifact of CNPs in lymphoblastoid cell line DNA. Because of the concern of potential in vitro chromosomal copy number changes in the course of cell culture of LCLs, we attempted to confirm the above deletions and insertions within a matched blood DNA sample from the same subject. At the same 9q31.1 and 9q21.2 regions where respective deletions and gains had been observed in LCL DNA, no evidence for a deletion was observed in the blood DNA (Fig 2e). For

9q31.1, the median CN of these 65 SNPs was 1.85, with a range of 1.6 to 2.1. There were 12 heterozygous among these 65 SNPs, compared with no heterozygous when these SNPs were assayed in LCL DNA. At the 9q21.2 region, no evidence of a gain was observed in the blood DNA (Fig 2f), and the median CN was 1.9 with a range of 1.7 to 2.0. Consistent with these results, qPCR assays failed to detect the deletion and insertion at these regions in blood DNA (Fig 2g-h). These data suggest the large-scale deletion and insertion observed in the LCL DNA of this subject were somatic changes that occurred in the course of cell culture. In addition, the average successful SNP call rates that we obtained were 99.20% from blood DNA, in contrast to 98.68% for LCL DNA. Therefore, it appears that DNA isolated from blood is more reliable than LCL DNA for studies of germline CNPs.

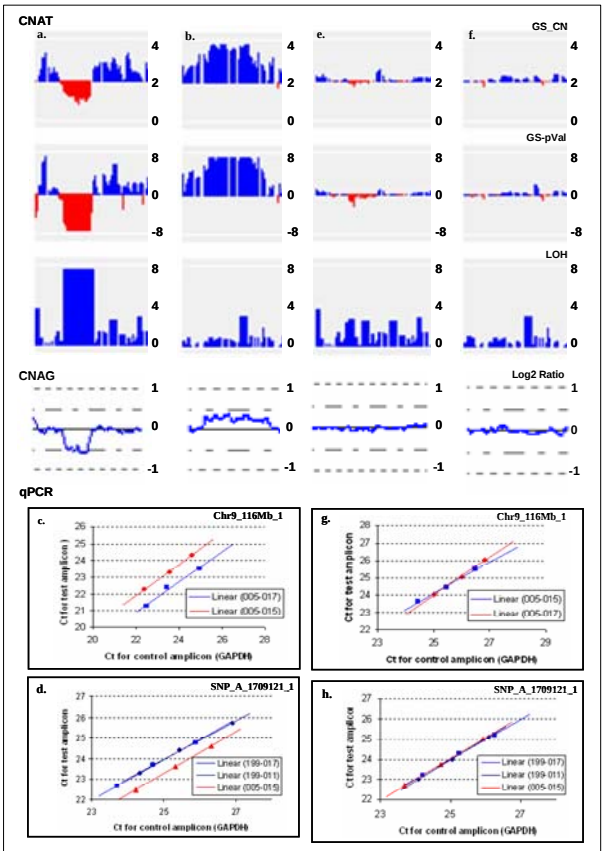

Fig 2. DNA copy number (CN) changes detected by 100K SNP mapping panel of the Affymetrix GeneChip®. A deletion at 9q31.1 (a) and a gain at 9q21.2 (b) were observed in lymphoblastoid cell line (LCL) DNA of a subject. Quantitative real-time PCR (qPCR) confirmed the deletion (c) and gain (d) in this sample. No deletion or gain was found in a blood DNA sample from the same individual, as measured either by the Gene Chip (e and f, respectively), or qPCR (g and h, respectively). In panels c, d, g, and h, the X-axis shows the averaged Ct values for the control (GAPDH) amplicon and the Y-axis shows the test amplicons for the three dilutions of each tested DNA sample (20, 10, 5ng). The Ct values for control and test amplicons at each dilution of DNA were plotted against each other, and the offset between best-fit lines along the test-amplicon axis indicates CN differences between samples. In the Gene Chip analysis, a CN difference of 1 between samples was found in LCL DNA (c and d) and no change was found in blood DNA (e and f).

CNPs >100 kb are rare in the genome. We screened CNPs in blood DNA from an additional 22 subjects in four HPC families with Affymetrix 100k SNP mapping sets. For putative deletions, we used the working criteria of a minimum of two consecutive SNPs with the following characteristics: CN < 1.2, P-value < 10-5.5, a separation of ≥ 0.55 in CN between groups of subjects, and homozygous genotypes. For putative gains, we used the working criteria of a minimum of two consecutive SNPs with the following characteristics: CN > 2.8, P-value < 10-4, and a separation of ≥ 0.55 in CN between groups of subjects. With these criteria, not a single region in the genome met the criteria for a deletion in any of the 23 subjects analyzed. However, we found four regions that met the criteria for gains, one of which is within a known replicon that is mapped to several chromosomes. For the remaining three regions, we performed qPCR and confirmed two of these gains (Table 1). One confirmed gain involved two SNPs spanning 32,790 bp at 10q11 and was found among multiple subjects in all three families. This CNP is also within a known replicon that is mapped to a single chromosomal region, and has been previously described (Sebat 2004; Iafrate 2004). The other confirmed gain was found in multiple subjects from a single family and involved four SNPs spanning 9,095 bp at 19q13. This is a novel CNP and is not within a known duplicon. The remaining region that was not confirmed by qPCR was found in a single subject. These results suggest that large-scale germline CNPs involving several hundred kb can be detected using the Affymetrix 100K SNP panel. However, the frequencies of large-scale germline CNPs in our study samples are not as common as predicted (Sebat 2004; Iafrate 2004).

CNPs involving 100s to 1,000s bp are more common. While the 500-kb smooth average is a good option for detecting large-scale CNPs, it may miss smaller CNPs because altered allele intensities of implicated SNPs may be averaged out by the allele intensities of SNPs in the flanking normal region. To test this, we searched the genome for small- and mid-scale CNPs by re-analyzing the data using smaller window sizes (100-kb and 10-kb). We found three regions that met the working criteria of deletions using a 100-kb window size, all of which were confirmed by qPCR (Table 1). One such deletion involved ten SNPs spanning 145,676 bp at 16q21. Using a 10-kb window size, we found nine additional regions that met the working criteria of deletions; five were confirmed by qPCR. For the remaining four regions, mutations were found in either probe or restriction enzyme sequences, which likely decreased the intensity of hybridization, thereby affecting CN calculation. We also found 18 regions that met the working criteria of gains using a 100-Kb window size, and three of the six regions selected for qPCR assays were confirmed. In addition, 50 additional regions met the working criteria of gains using a 10-Kb window size, and only one of the six regions examined by qPCR assays was confirmed.

Table 1. Germline CNPs identified by 100K SNP mapping panel of Affymetrix and confirmed by quantitative real-time PCR

Type of CNPs	Detection methods	Chromosoma region	Implicated SNPs #	Positions (bp)	Size (bp)	Multiple subjects	Repetative regions	Genic regions	Previously reported
Deletion	100 kb smooth window	2q12.1	2	105,169,462 - 105,170,804	1,342	Y	N	N	N
		16q21	10	63,032,479 - 63,178,155	145,676	N	N	N	N
		20p12.2	2	9,868317 - 9,893-310	24,993	N	N	N	N
	10 kb smooth window	2q32.1	2	185,000,369 - 185,002,549	2,180	Y	N	Y	N
		3p14.2	3	62,951,975 - 62,952,488	69,187	N	N	Y	N
		4q28.3	2	135,518,614 - 135,519,403	789	Y	N	N	N
		12q23.1	3	98,791,609 - 98,791,950	341	Y	N	Y	N
		12q24.23	2	118,401,599 - 118,401,998	399	Y	N	Y	N
Gain	500 kb smooth window	10q11.22	2	46,465,579 - 46,498,369	32,790	Y	Y	Y	Y
		19q13.41	4	57,024,188 - 57,033,283	9,095	Y	N	N	N
	100 kb smooth window	5p13.31	4	8,920,157 - 8,956,052	35,895	Y	N	Y	N
		14q11.2	2	18,205,576 - 18,218,052	12,476	Y	Y	Y	Y
		17p13.2	5	5,632,970 - 5,726,274	93,304	Y	N	Y	N
	10 kb smooth window	7p21.2	4	14,785,866 - 14,790,343	4,477	Y	N	N	N

Among the 14 confirmed CNPs in our study, 13 were novel. Two of six gains were located within replicons, but none of the eight deletions was within a replicon. Eight CNPs, including 4 deletions and 4 gains, involved either known genes or predicted genes. The size of the CNPs ranged from 341 bp to 145,676 bp.

Need for a higher resolution of SNP panel. Our data suggested that the Affymetrix 100K SNP mapping panel can be used to identify germline CNPs. However, we recognize that our study may miss a substantial number of smaller CNPs due to limited SNP resolution. This potential limitation can be overcome by using the higher resolution 500K SNP panel from Affymetrix, which has a median inter-SNP distance of 2.5 kb.

Identification of CNPs using the Affymetrix 500K SNP Mapping Panel. We recently tested the Affymetrix 500K in identifying germline CNPs. The results are very encouraging and support that we can use the panel to accurately identify germline CNPs.

Identifying germline CNPs among 48 HapMap subjects using Affymetrix 500K mapping panel. To assess the feasibility of using the Affymetrix 500k SNP mapping set to detect CNPs, we analyzed an Affymetrix 500k SNP mapping dataset, which contains allele intensity data for 48 individuals (including 5 HapMap CEPH trios, 5 Yoruban trios, three other non-HapMap trios, and 9

unrelated HapMap Asian samples). For putative deletions, we used the working criteria of a minimum of three out of four consecutive SNPs with a separation of ≥0.5 in CN between groups of subjects, and homozygous genotypes. For putative gains, we used the working criteria of a minimum of three out of four consecutive SNPs with a separation of ≥0.5 in CN between groups of subjects. A 1-kb smooth window was used to generate copy number. The criteria we used here is slightly different from the ones we used previously for the 100K mapping panel because of different computer programs (dChip vs. CNAT). We only included potential CNPs that 1) appear at least two times in these 48 individuals, and 2) follow Medelian inheritance in trios. These CNPs have higher likelihood to be true CNPs. Alltogether, we identified 824 deletions and 722 amplifications. The average, median, and size ranges for these CNPs are summarized in Table 5. Some of these CNPs are relatively common. Fifty deletions and forty-four amplifications identified have frequencies > 5%. There are 337 deletions and 332 amplifications in the genic region. Sixteen deletions and 28 amplifications involve at least one exon (Table 2). Interestingly, we observed several ethnicity-specific CNPs.

Table 2. Summary for CNPs identified in 48 individuals using Affymetrix 500k SNP mapping panel

Type of CNPs	# of CNPs identified	Average size (kb)	Median size (kb)	Size range (kb)	# of CNPs with allele frequency > 0.05	# of CNPs in genic regions	# of CNPs covering exonic regions
Deletion	824	47.4	18.2	1-2,395	50	337	16
Amplification	722	53.1	24.7	1-3,204	44	332	28

We identified germline CNPs among four PCa probands in our HPC families using the Affymetrix 500K mapping panel. The average call rates were 93.32% for the Sty array and 95.63% for the Nsp array, respectively. The same criteria used to identify deletions and amplifications as mentioned above were applied. Altogether, we identified 35 deletions and 109 amplifications. Within these CNPs, 13 deletions and 49 amplifications are in genic regions. Recently, we completed genotyping of Affymetrix 500K SNP arrays among 206 HPC probands.

A deletion involving a candidate PCa tumor suppressor gene. Interestingly, one of the identified genic deletions involves a deletion of the last 4 exons of a tumor suppressor gene WWOX (WW domain-containing oxidoreductase isoform 2). This gene encodes a protein which contains 2 WW domains and a short-chain dehydrogenase/reductase domain (SRD). The highest normal expression of this gene is detected in hormonally regulated tissues such as testis, ovary, and prostate. This expression pattern and the presence of an SRD domain suggest a role for this gene in steroid metabolism. In addition, it was also implicated in tumor necrosis factor (TNF)-mediated cell death, as well as p53 controlled genotoxic stress-induced cell death. Loss of heterozygosity and down-regulation of this tumor suppressor gene in PCa has been reported[11]. Homozygous deletion of WWOW exons has been reported in ovarian cancer[12], and hepatocellular carcinoma (Yakicier 2001). Intriguingly, this particular CNP is rare in the 48 HapMap subjects (allele frequency =1.4%; the mother and son of one trio carry this CNP), but two out of the four PCa subjects we genotyped carry this CNP. It is likely that germline deletions of a tumor suppressor gene (which would be undetected in a regular LOH study), followed by somatic deletions or mutations of the wild-type allele, contributes to tumorigenesis.

Key Research Accomplishments

Published one paper that is directly related to this grant.

Liu W, Chang B, Li T, Dimitrov L, Kim S, Kim JW, Turner AR, Meyers DA, Trent JM, Zheng SL, Isaacs WB, Xu J. Germline copy number polymorphisms involving larger than 100 kb are uncommon in normal subjects. Prostate. 2007 Feb 15;67(3):227-33.

Published several papers that are related to this grant.

Zheng SL, Sun J, Cheng Y, Li G, Hsu FC, Zhu Y, Chang BL, Liu W, Kim JW, Turner AR, Gielzak M, Yan G, Isaacs SD, Wiley KE, Sauvageot J, Chen HS, Gurganus R, Mangold LA, Trock BJ, Gronberg H, Duggan D, Carpten JD, Partin AW, Walsh PC, Xu J, Isaacs WB. Association between two unlinked loci at 8q24 and prostate cancer risk among European Americans. J Natl Cancer Inst. 2007 Oct 17;99(20):1525-33. Epub 2007 Oct 9.

Liu W, Chang BL, Cramer S, Koty PP, Li T, Sun J, Turner AR, Von Kap-Herr C, Bobby P, Rao J, Zheng SL, Isaacs WB, Xu J. Deletion of a small consensus region at 6q15, including the MAP3K7 gene, is significantly associated with high-grade prostate cancers. Clin Cancer Res. 2007 Sep 1;13(17):5028-33.

Liu W, Ewing CM, Chang BL, Li T, Sun J, Turner AR, Dimitrov L, Zhu Y, Sun J, Kim JW, Zheng SL, Isaacs WB, Xu J. Multiple genomic alterations on 21q22 predict various TMPRSS2/ERG fusion transcripts in human prostate cancers. Genes Chromosomes Cancer. 2007 Nov;46(11):972-80.

Chang BL, Liu W, Sun J, Dimitrov L, Li T, Turner AR, Zheng SL, Isaacs WB, Xu J. Integration of somatic deletion analysis of prostate cancers and germline linkage analysis of prostate cancer families reveals two small consensus regions for prostate cancer genes at 8p. Cancer Res. 2007 May 1;67(9):4098-103.

Sun J, Liu W, Adams TS, Sun J, Li X, Turner AR, Chang B, Kim JW, Zheng SL, Isaacs WB, Xu J. DNA copy number alterations in prostate cancers: a combined analysis of published CGH studies. Prostate. 2007 May 15;67(7):692-700.

Liu W, Chang B, Sauvageot J, Dimitrov L, Gielzak M, Li T, Yan G, Sun J, Sun J, Adams TS, Turner AR, Kim JW, Meyers DA, Zheng SL, Isaacs WB, Xu J. Comprehensive assessment of DNA copy number alterations in human prostate cancers using Affymetrix 100K SNP mapping array. Genes Chromosomes Cancer. 2006 Nov;45(11):1018-32.

Reportable Outcomes

1) Discovered germline CNPs in the genome by measuring the allele intensity of 500K SNPs.

2) Most of germline CNPs are inherited from parents.

3) Majority of detected CNPs can be confirmed using real-time PCR.

Conclusion

This novel and systematic approach, when applied to this high-risk hereditary study population, increases the likelihood to identify important genetic alterations that predispose to prostate cancer risk. To our knowledge, our study is the first of its kind. If our study is successful, it will likely contribute to our understanding of prostate cancer etiology, and provide novel targets for prostate cancer risk assessment, prevention, and therapy.

References

Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, Chi M, Navin N, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A, Wigler M (2004). Large-scale copy number polymorphism in the human genome. Science 305(5683):525-528.

Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C. (2004) Detection of large-scale variation in the human genome. Nat Genet. 36:949-51.

Conrad DF, Andrews TD, Carter NP, Hurles ME, Pritchard JK. (2006) A high-resolution survey of deletion polymorphism in the human genome. Nat Genet. 38:75-81.

Hinds DA, Kloek AP, Jen M, Chen X, Frazer KA. (2006) Common deletions and SNPs are in linkage disequilibrium in the human genome. Nat Genet. 38:82-5.

McCarroll SA, Hadnott TN, Perry GH, Sabeti PC, Zody MC, Barrett JC, Dallaire S, Gabriel SB, Lee C, Daly MJ, Altshuler DM; International HapMap Consortium. (2006) Common deletion polymorphisms in the human genome. Nat Genet. 38:86-92.

Yakicier MC, Legoix P, Vaury C, Gressin L, Tubacher E, Capron F, Bayer J, Degott C, Balabaud C, Zucman-Rossi J. (2001) Identification of homozygous deletions at chromosome 16q23 in aflatoxin B1 exposed hepatocellular carcinoma. Oncogene. 20:5232-8

Appendices.

Liu W, Chang B, Li T, Dimitrov L, Kim S, Kim JW, Turner AR, Meyers DA, Trent JM, Zheng SL, Isaacs WB, Xu J. Germline copy number polymorphisms involving larger than 100 kb are uncommon in normal subjects. Prostate. 2007 Feb 15;67(3):227-33.

Liu W, Chang BL, Cramer S, Koty PP, Li T, Sun J, Turner AR, Von Kap-Herr C, Bobby P, Rao J, Zheng SL, Isaacs WB, Xu J. Deletion of a small consensus region at 6q15, including the MAP3K7 gene, is significantly associated with high-grade prostate cancers. Clin Cancer Res. 2007 Sep 1;13(17):5028-33.

Liu W, Ewing CM, Chang BL, Li T, Sun J, Turner AR, Dimitrov L, Zhu Y, Sun J, Kim JW, Zheng SL, Isaacs WB, Xu J. Multiple genomic alterations on 21q22 predict various TMPRSS2/ERG fusion transcripts in human prostate cancers. Genes Chromosomes Cancer. 2007 Nov;46(11):972-80.

Chang BL, Liu W, Sun J, Dimitrov L, Li T, Turner AR, Zheng SL, Isaacs WB, Xu J. Integration of somatic deletion analysis of prostate cancers and germline linkage analysis of prostate cancer families reveals two small consensus regions for prostate cancer genes at 8p. Cancer Res. 2007 May 1;67(9):4098-103.

Sun J, Liu W, Adams TS, Sun J, Li X, Turner AR, Chang B, Kim JW, Zheng SL, Isaacs WB, Xu J. DNA copy number alterations in prostate cancers: a combined analysis of published CGH studies. Prostate. 2007 May 15;67(7):692-700.

Liu W, Chang B, Sauvageot J, Dimitrov L, Gielzak M, Li T, Yan G, Sun J, Sun J, Adams TS, Turner AR, Kim JW, Meyers DA, Zheng SL, Isaacs WB, Xu J. Comprehensive assessment of DNA copy number alterations in human prostate cancers using Affymetrix 100K SNP mapping array. Genes Chromosomes Cancer. 2006 Nov;45(11):1018-32.

The Prostate 67:227–233 (2007)

Germline Copy Number Polymorphisms Involving Larger Than 100 kb Are Uncommon in Normal Subjects

Wennuan Liu,[1] Baoli Chang,[1] Tao Li,[1] Latchezar Dimitrov,[1] Seungchan Kim,[2,3] Jin Woo Kim,[1] Aubrey R. Turner,[1] Deborah A. Meyers,[1] Jeffery M. Trent,[2] Siqun Lilly Zheng,[1] William B. Isaacs,[4] and Jianfeng Xu[1,2]*

[1]*Center for Human Genomics, Wake Forest University School of Medicine, Winston-Salem, North Carolina*
[2]*Translational Genomics Research Institute (TGen), Phoenix, Arizona*
[3]*Department of Computer Science and Engineering, Ira A. Fulton School of Engineering, Arizona State University, Tempe, Arizona*
[4]*Department of Urology, Johns Hopkins Medical Institutions, Baltimore, Maryland*

BACKGROUND. Recent studies using ROMA and Array-CGH suggest that germline copy number polymorphisms (CNPs) involving >100 kb are common in humans.
METHODS. In this study, we used the Affymetrix GeneChip 100K single nucleotide polymorphisms (SNP) mapping panel to further examine the type and frequency of germline CNPs in the genome. By utilizing the allele intensity data generated while genotyping ~116,000 SNPs among 23 subjects from 4 families, we were able to detect multiple CNPs.
RESULTS. However, in contrast to several previous studies, we found that CNPs >100 kb are rare in the genome but CNPs involving 100s–1,000s of base pairs are more common.
CONCLUSIONS. We have demonstrated the utility of this approach, which has an important advantage over other methods because it is able to simultaneously assess both CNPs and SNPs, and therefore has great potential in genetic association studies of common diseases. *Prostate 67: 227–233, 2007.* © 2006 Wiley-Liss, Inc.

KEY WORDS: **germline; DNA copy number polymorphisms (CNPs); genomewide**

INTRODUCTION

Using a representational oligonucleotide microarray analysis (ROMA), Sebat et al. [1] performed a genome-wide analysis of germline gross deletions/gains. Among 20 normal subjects, they found 221 DNA copy number differences representing 76 unique copy number polymorphisms (CNPs) in the genome, with a median length of 222 kb. Similarly, Iafrate et al. [2] reported 255 CNPs in the genome among 39 unrelated healthy individuals and 16 individuals with known chromosomal imbalances using array-based comparative genomic hybridization (array-CGH). Together with several other reports [3–5], it appears germline CNPs in the genome are more common than previously estimated.

To further understand germline CNPs in the genome, we utilized the 100K single nucleotide polymorphisms (SNP) mapping panel of the Affymetrix GeneChip for systematic detection of germline CNPs. This alternative is appealing because it allows for

Wennuan Liu and Baoli Chang contributed equally to the study.

Grant sponsor: National Institutes of Health; Grant numbers: CA106523, CA95052; Grant sponsor: Department of Defense; Grant number: PC051264.

*Correspondence to: Dr. Jianfeng Xu, Medical Center Blvd, Winston-Salem, NC 27157. E-mail: jxu@wfubmc.edu

Received 24 January 2006; Accepted 13 February 2006
DOI 10.1002/pros.20441
Published online 27 December 2006 in Wiley InterScience (www.interscience.wiley.com).

© 2006 Wiley-Liss, Inc.

WILEY InterScience® DISCOVER SOMETHING GREAT

measurement of both allele intensity and genotypes of SNPs, and this combination provides information, that is, important in defining CNPs, especially deletions.

MATERIALS AND METHODS

Microarray Analysis

Four families, including 23 subjects with or without prostate cancer and women, were selected from a large set of 188 hereditary prostate cancer families collected at Johns Hopkins Hospital [6]. Genomic DNA from blood and lymphoblastoid cell lines (LCL) was assayed using the 100K SNP mapping panel following the manufacturer's standard protocol. Briefly, 250 ng of genomic DNA was digested with either Hind III or Xba I, and then ligated to adapters that recognize the cohesive four basepair (bp) overhangs. A generic primer that recognizes the adapter sequence was used to amplify adapter-ligated DNA fragments with PCR conditions optimized to preferentially amplify fragments in the 250–2,000 bp size range in a GeneAmp PCR System 9700 (Applied Biosystems, Foster City, CA). After purification with a QIAGEN MinElute 96 UF PCR purification system, a total of 40 μg of PCR product was fragmented and about 2.9 μg was visualized on a 4% TBE agarose gel to confirm that the average size was smaller than 180 bp. The fragmented DNA was then labeled with biotin and hybridized to the GeneChip Mapping 100K Set for 17 hr at 48°C in a Hybridization Oven 640. We washed and stained the arrays using Affymetrix Fluidics Station 450 and scanned the arrays using a GeneChip Scanner 3000 G7 (Affymetrix, Inc., Santa Clara, CA).

The Affymetrix GeneChip Operating Software (GCOS) collected and extracted feature data from Affymetrix GeneChip Scanners. We used the GeneChip DNA analysis software (GDAS) to analyze cell intensity data stored in the GCOS Database for genotyping using a Dynamic Model mapping algorithm. We calculated DNA copy numbers using Chromosome Copy Number Analysis Tool (CNAT) version 2.0, Copy Number Analyzer for Affymetrix GeneChip (CNAG) [7], and dChipSNP [8]. Gaussian kernel-smoothing average was used for averaging the copy number and *P*-value of individual SNPs over a fixed genomic interval (10, 100, or 500 kb window size). The smoothing averages out the random noise across flanking SNPs and minimizes the false-positive rate, while keeping the true-positive rate high. The kernel-smoothing accentuates genomic intervals in which consecutive SNPs display the same type of alteration (gain or loss). The 100K SNP mapping panel contains 116,204 SNP probes, with a median physical distance between SNPs of 8.5 kb. The average successful SNP call rates were 99.20 and 98.68% for blood and LCL DNA, respectively.

Confirmation via qPCR

A subset of putative CN changes was subjected to confirmation by quantitative real-time PCR (qPCR) using the ABI Prism 7000 Sequence Detection System, and direct sequencing using the ABI 3700 DNA Analyzer. For qPCR, primers were designed using Primer Express 1.5 software from Applied Biosystems. Amplicons were designed against the putatively hemizygous locus and a control locus of known normal copy number. The PCR kinetics at the control locus was used to control for sample-to-sample differences in genomic DNA purity and concentration. Three concentrations of each genomic DNA sample (20, 10, and 5 ng) were assayed in duplicate, using each pair of real-time-PCR primers. PCRs were prepared as follows: in 20 μl, we combined 2 μl of genomic DNA, 0.05 μM of each primer, and SYBR-Green PCR Master Mix from Applied Biosystems. PCRs were performed as follows: 95°C for 10 min, followed by 40 cycles at 95°C for 20 sec, and 60°C for 1 min. An additional cycle of 95°C for 15 sec, 60°C for 20 sec, and 95°C for 15 sec was run at the end to measure the dissociation curve for qualify control. We used the Sequence Detection Software (SDS) for PCR baseline subtraction and exported the threshold cycle number (Ct) data for analysis. Ct values for the control and test amplicons for the three dilutions of each DNA sample were plotted against each other, and the offset between two samples along the control-amplicon axis and test-amplicon axis was measured. An offset of 0.8–1.2 along the test-amplicon axis was taken to indicate a copy number difference of 1 to 2 ratios between the two samples at that locus. For direct sequencing, we designed PCR primers to cover the entire regions of the Hind III or Xba I restriction fragments using primer 3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) and synthesized by Integrated DNA Technologies, Inc (Coralville, IA). We used Platinum *Pfx* DNA polymerase from Invitrogen to produce PCR products for sequencing. All PCR products were purified using the QIAquick PCR purification Kit (Qiagen) to remove deoxynucleoside triphosphates and excess primers. We performed all sequencing reactions using dye-terminator chemistry (BigDye; ABI, Foster City, CA), and separated the products in an ABI 3700 DNA Analyzer. We identified SNPs and other small changes using Sequencher software version 4.0.5 (Gene Codes Corporation).

RESULTS AND DISCUSSION

We began with an LCL DNA sample from a prostate cancer patient (005–015). Allele intensities of over 100K SNPs were analyzed using default settings of CNAT (500-kb smooth averages) and CNAG. Two large-scale deletions and four large-scale gains were evident in

chromosome 9 (Fig. 1). For example, at the 9q31.1 region, 65 consecutive SNPs spanning 2.4 Mb signaled a deletion: CN $\leq$ 1.2 and *P*-values $\leq 10^{-5.5}$ (Fig. 2a). Similarly, at 9q21.2, a set of 24 consecutive SNPs spanning 0.5 Mb revealed a gain: CN $\geq$ 2.8 and *P*-values $\leq 10^{-4}$ (Fig. 2b). The deletion and gain were confirmed by qPCR (Fig. 2c,d). These data suggest that the Affymetrix 100K SNP panel is able to detect large-scale deletions and insertions. Furthermore, these data provide an empirical basis for defining the criteria that may be used to define deletions and insertions.

Because of the potential concern of in vitro chromosomal copy number changes in the course of cell culture of LCLs, we attempted to confirm the above deletions and insertions within a matched blood DNA sample from the same subject using the Xba chip of the 100K SNP panel. At the same 9q31.1 and 9q21.2 regions where respective deletions and gains had been observed in LCL DNA, no evidence for a deletion was observed in the blood DNA (Fig. 2e). For 9q31.1, the median CN of these 65 SNPs was 1.85, with a range of 1.6–2.1. There were 12 heterozygous among these 65 SNPs, compared with no heterozygous when these SNPs were assayed in LCL DNA. At the 9q21.2 region, no evidence of a gain was observed in the blood DNA (Fig. 2f), and the median CN was 1.9 with a range of 1.7–2.0. Consistent with these results, quantitative real-time PCR assays failed to detect the deletion and insertion at these regions in blood DNA (Fig. 2g,h). These data suggest the large-scale deletion and insertion observed in the LCL DNA of this subject were somatic changes that occurred in the course of cell culture. Therefore, it appears that DNA isolated from blood is more reliable than LCL DNA for studies of germline CNPs.

We therefore used blood DNA for the remaining analyses. Our primary goal was to detect large-scale CNPs involving >100 kb in the genome using CNAT analyses with a 500-kb window size. We screened for putative deletions using the working criteria of a minimum of two consecutive SNPs with the following characteristics: CN < 1.2, *P*-value $<10^{-5.5}$, a separation of $\geq$0.55 in CN between groups of subjects, and homozygous genotypes. We screened for putative gains using the working criteria of a minimum of two consecutive SNPs with the following characteristics: CN > 2.8, *P*-value $<10^{-4}$, and a separation of $\geq$ 0.55 in CN between groups of subjects. With these criteria, not-

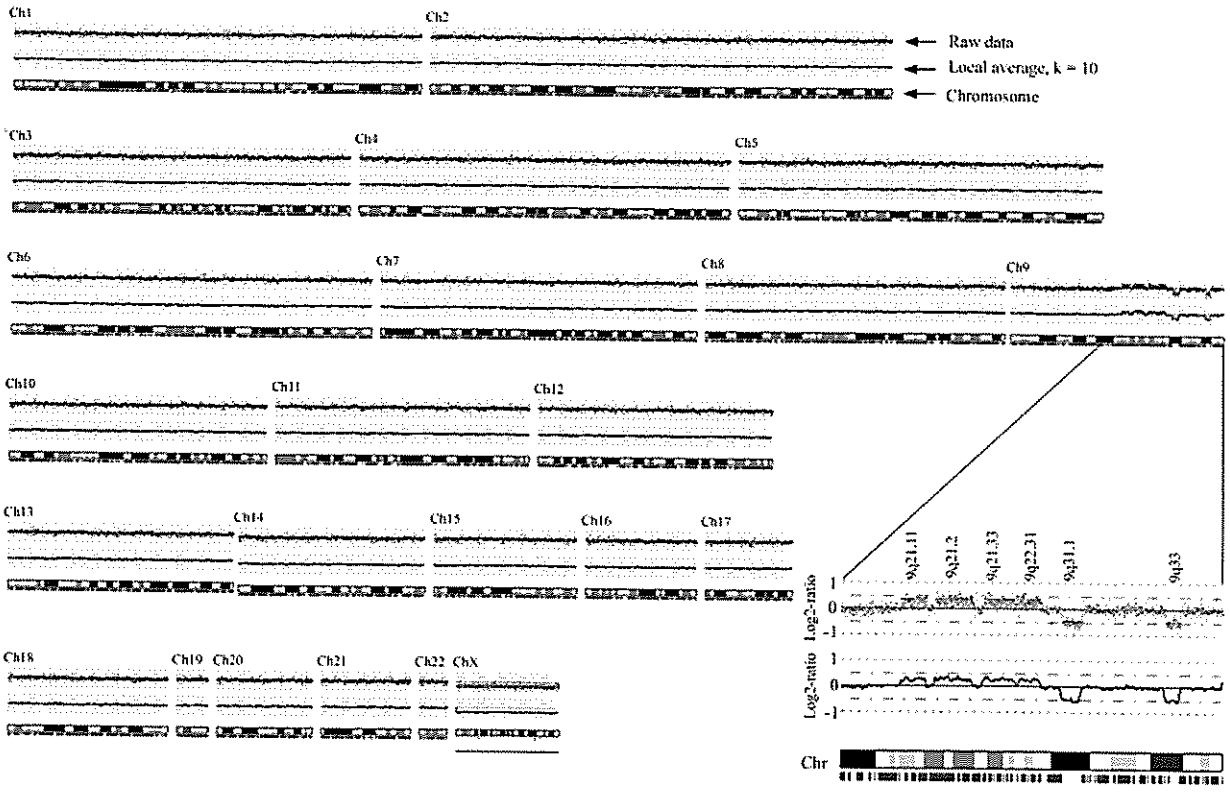

Fig. 1. Genome-wide copy number analysis of GeneChip 50K Xba array data of 005 – 015 lymphoblastoid cell line (LCL) DNA using CNAG. Chr, chromosome. Red dots, raw log2 ratio for each SNP. Green line, inferred copy number from a hidden Markov model. Blue curves, local mean analysis of ten consecutive SNPs. Green bar, heterozygous SNPs. Green bar, heterozygous SNP calls. LOH is marked under chromosome with blue lines.

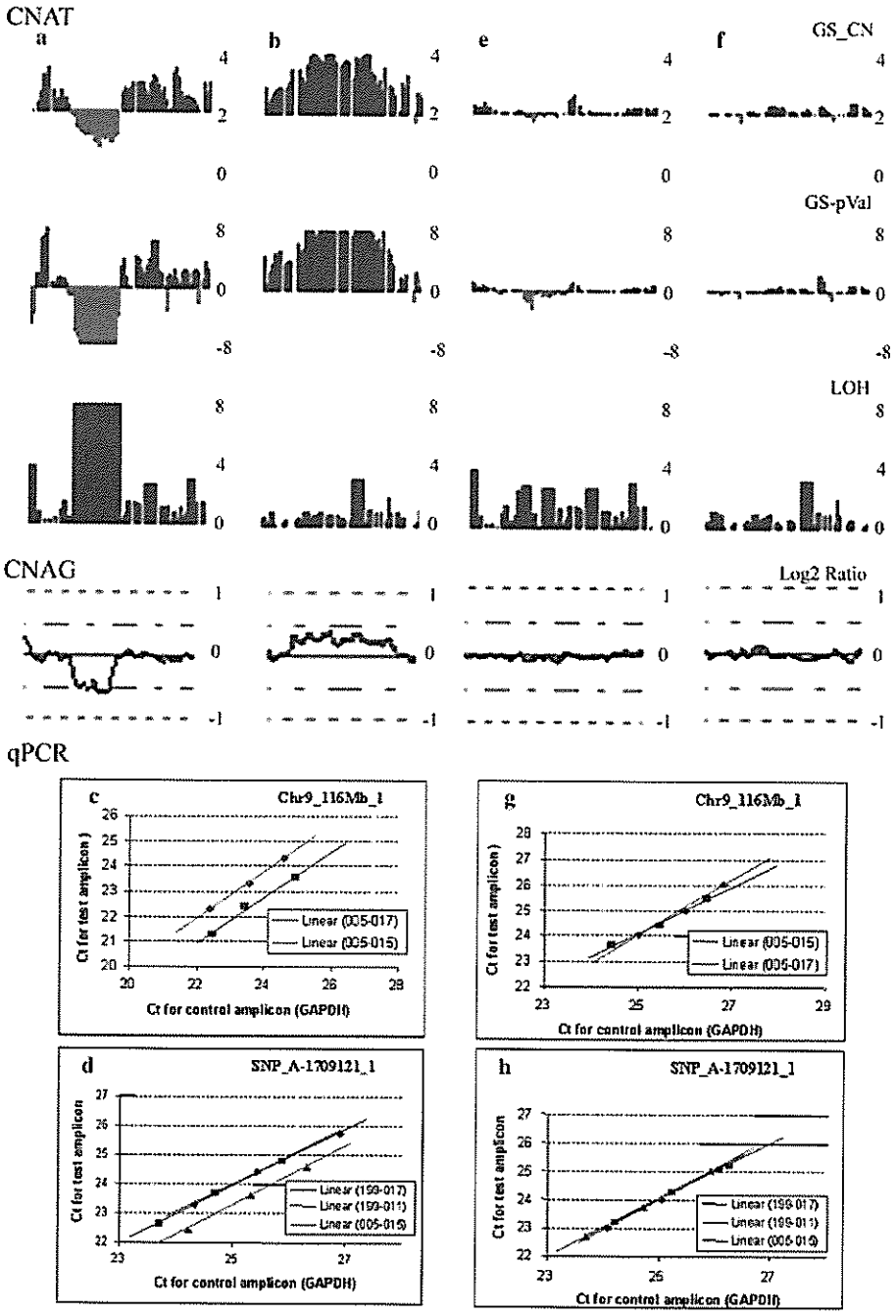

Fig. 2. DNA copy number (CN) changes detected by 100K SNP mapping panel of the Affymetrix GeneChip®. A deletion at 9q31.1 (**a**) and a gain at 9q21.2 (**b**) were observed in LCL DNA of a subject. Quantitative real-time PCR (qPCR) confirmed the deletion (**c**) and gain (**d**) in this sample. No deletion or gain was found in a blood DNA sample from the same individual, as measured either by the Gene Chip (**e** and **f**, respectively), or qPCR (**g** and **h**, respectively). In panels c, d, g, and h, the X-axis shows the averaged Ct values for the control (GAPDH) amplicon and the Y-axis shows the test amplicons for the three dilutions of each tested DNA sample (20, 10, 5 ng). The Ct values for control and test amplicons at each dilution of DNA were plotted against each other, and the offset between best-fit lines along the test-amplicon axis indicates CN differences between samples. In the Gene Chip analysis, a CN difference of one between samples was found in LCL DNA (c and d) and no change was found in blood DNA (e and f).

a single region in the genome met the criteria for a deletion for any of the 23 subjects analyzed. However, we found four regions that met the criteria for gains, one of which is within a known replicon that is mapped to several chromosomes. For the remaining three regions, we performed qPCR and confirmed two of these gains (Table I). One confirmed gain involved two SNPs spanning 32,790 bp at 10q11 (46,465,579–46,498,369 bp) and was found among multiple subjects in all three families. This CNP is also within a known replicon that is mapped to a single chromosomal region, and has been previously described [1,2]. The other confirmed gain was found in multiple subjects from a single family and involved four SNPs spanning 9,095 bp at 19q13 (57,024,188–57,033,283 bp). This is a novel CNP and is not within a known duplicon. The remaining region that was not confirmed by qPCR was found in a single subject. These results suggest that large-scale germline CNPs involving several hundred kbs can be detected using the Affymetrix 100K SNP panel. However, the frequencies of large-scale germline CNPs in our study samples are not as common as predicted by Sebat et al. [1] and Iafrate et al. [2].

While the 500-kb smooth average is a good option for detecting large-scale CNPs, it may miss smaller CNPs because altered allele intensities of implicated SNPs may be averaged out by the allele intensities of SNPs in the flanking normal region. We therefore searched the genome for small- and mid-scale CNPs by re-analyzing the data using smaller window sizes (100- and 10-kb). We found three regions that met the working criteria of deletions using a 100-kb window size, all of which were confirmed by qPCR (Table I). One such deletion involved 10 SNPs spanning 145,676 bp at 16q21. Using a 10-kb window size, we found nine additional regions that met the working criteria of deletions; five were confirmed by qPCR (Table I). For the remaining four regions, mutations were found in either probe or restriction enzyme sequences, which likely decreased the intensity of hybridization thereby affecting CN calculation. We also found 18 regions that met the working criteria of gains using a 100-Kb window size, and 3 of the 6 regions selected for qPCR assays were confirmed (Table I). In addition, 50 additional regions met the working criteria of gains using a 10-Kb window size, and only 1 of the 6 regions examined by qPCR assays was confirmed (Table I).

Among the 14 confirmed CNPs in our study, 13 were novel. Two of six gains were located within replicons, but none of the eight deletions was within a replicon. Eight CNPs, including four deletions and four gains involved either known genes or predicted genes. The size of the CNPs ranged from 341 to 145,676 bp.

All together, we observed numerous germline CN changes in the genome, the majority of which were

TABLE I. Germline CNPs Identified by 100K SNP Mapping Panel of Affymetrix and Confirmed by Quantitative Real-Time PCR

Type of CNPs	Detection methods	Chromosomal region	Implicated SNPs #	Implicated SNPs Positions (bp)	Size (bp)	Multiple subjects	Repetative regions	Genic regions	Previously reported
Deletion	100-kb smooth window	2q12.1	2	105,169,462–105,170,804	1,342	Y	N	N	N
		16q21	10	63,032,479–63,178,155	145,676	N	N	N	N
		20p12.2	2	9,868,317–9,893-310	24,993	N	N	N	N
	10-kb smooth window	2q32.1	2	185,000,369–185,002,549	2,180	Y	N	Y	N
		3p14.2	3	62,951,975–62,952,488	69,187	N	N	Y	N
		4q28.3	2	135,518,614–135,519,403	789	Y	N	N	N
		12q23.1	3	98,791,609–98,791,950	341	Y	N	Y	N
		12q24.23	2	118,401,599–118,401,998	399	Y	N	Y	N
Gain	500-kb smooth window	10q11.22	2	46,465,579–46,498,369	32,790	Y	Y	Y	Y
		19q13.41	4	57,024,188–57,033,283	9,095	Y	N	N	N
	100-kb smooth window	5p13.31	4	8,920,157–8,956,052	35,895	Y	N	Y	N
		14q11.2	2	18,205,576–18,218,052	12,476	Y	Y	Y	Y
		17p13.2	5	5,632,970–5,726,274	93,304	Y	N	Y	N
	10-kb smooth window	7p21.2	4	14,785,866–14,790,343	4,477	Y	N	N	N

small (100s–1,000s bp). A considerable difference of our findings compared with the previous studies was that CNPs >100 kb were rare. A combination of factors may account for the difference. First, the sample sizes and characteristics of study subjects differ. The Sebat study [1] included 20 unrelated healthy subjects, and the Iafrate study [2] included 39 unrelated healthy subjects and 16 subjects with known chromosomal imbalances. Our study included 23 subjects from four prostate cancer families. These 23 subjects represent 8 independent founders. Based on the estimated frequency of large-scale CNPs in the genome [1], though, we would expect to observe 77 CNPs of ~222 kb in our study. Second, the genome coverage of the SNPs and probes differs. However, the 100K SNP panel provides similar coverage of the genome and of previously reported CNPs regions. For example, among the 76 unique CNPs identified by Sebat [1], 51 CNPs are covered by at least two SNPs, and the majority are covered by >10 SNPs. Third, the source of genomic DNA may affect the frequency of CNPs. Sebat et al. [1] examined genomic DNA from LCLs, Iafrate et al. [2] examined DNA from both LCLs and blood, and we examined DNA from blood. Artifact CNPs may arise in cell lines during the steps of generation, manipulation, and culturing. This possibility was demonstrated by the comparison of CNPs between the matched DNA samples isolated from blood and LCL of the same individual (005–015) in our study. It is unclear whether the different results of LCL and blood DNA are specific to this cell line. The potential for artifacts that may appear to be CNPs should be carefully considered whenever cell lines are assayed for the presence of CNPs. Fourth, the choice of reference samples differ. Our use of a large number of subjects (N = 110) as a reference group may lead to more stable results compared to a single subject reference in both ROMA and BAC array-CGH. Finally, our utilization of genotype information (e.g., heterozygous results), and our efforts to confirm many of the putative CNPs by qPCR and direct sequencing in this study may guard against false positive deletions.

The frequencies of copy number gains and losses are dependent on the degree of genetic diversity within the population used for the study and on the resolution of the microarray platform. The genetic diversity within our study was limited because the study population was composed of 23 Caucasian subjects from 4 families. The 100K SNPs microarray used in this study has better resolution than most of the previous studies; however, as the resolution of microarrays increase in future studies, the frequency of identifiable CNPs in humans will certainly increase. Therefore, it is premature to estimate the frequency of CNPs in the human genome based upon this study. It is more appropriate to summarize by stating that we have demonstrated smaller CNPs were much more frequent than large CNPs. The results of our study can best contribute to a more accurate estimate of CNPs in the human genome as our methods are applied to additional study populations, using microarrays of increasing resolution.

There are several reasons that many of the CNPs detected by SNP arrays could not be confirmed by qPCR. First, the algorithm used by the software to estimate DNA copy number affects the accuracy of the copy number calculation, which is beyond the scope of this work. Second, the conditions of PCR-based preparation of DNA for microarray hybridization might randomly cause artificial imbalances in different sequences of the genome. Third, mutations in the sequences of the SNP probes and restriction enzymes are the most obvious factors that may affect the intensity of hybridization, thereby affecting the accuracy of the copy number calculation. Overall, this means all potential CNPs should be verified by additional molecular approaches before they are considered confirmed.

While we were preparing our manuscript, three additional studies examining germline deletions based on SNP information in the human population were published [9–11]. In addition to identifying a large number of germline deletions in the genome, their results also suggest that the germline deletions, with median lengths ranging from 500 bp to 10.5 kb, are smaller than the studies of Sebat [1], Iafrate [2], and Tuzun [3]. Although the conclusions regarding germline deletions are similar between these three SNP-based studies and our study, there are at least two major differences. One is the methods used to identify CNPs; the studies by Conrad [9] and McCarroll [11] relied on family-based methods, which are not applicable to many case-control association studies. In contrast, our method utilized the allele intensity data generated from SNP genotyping and can be readily applied to both family-based and case-control studies. This is an important issue because Affymetrix SNP genotyping intensity data is currently being generated for qualitative genotyping as part of several genome-wide SNP association studies currently underway. Another difference is that our method can identify both germline deletions and gains, while the other three recent studies focused only on germline deletions. In fact, we found that there are more gains than deletions in our study population. The unique ability of our method to examine both deletions and gains allowed us to fully test the claims of the two initial reports [1,2] that large-scale germline CNPs involving ≥100 kb are common in the genome.

CONCLUSIONS

We have shown that by combining data on the allele intensity and genotype of SNPs, we can detect CNPs in the genome. The fact that a substantial number of detected CN changes, especially deletions, can be confirmed by qPCR and direct sequencing, as well as the fact that the same CNPs were detected in multiple individuals within a family, validates this method. If CNPs of >100 kb are frequent in the genome, this method should be able to detect at least a large subset of these large-scale CNPs. On the other hand, we recognize that our study may miss a substantial number of smaller CNPs due to limited SNP resolution. This potential limitation can be overcome by using the higher resolution 500K SNP panel from Affymetrix.

Based on the observed sizes and frequencies of CNPs, a major implication of our study is that future examinations of germline CNPs should focus on smaller-scale deletions and gains. There is a growing list of genes that seem to intersect with the sites of deletion. Conrad and colleagues [9], for example, identified 92 genes that were entirely deleted and another 109 genes in which coding sequences were partially eliminated. These types of CNPs, together with SNPs which are present at much higher frequencies may affect the function and expression of disease risk and modifier genes. The SNP mapping panel offers a unique possibility to simultaneously assess the roles of SNPs and CNPs in disease risk [12], although it is premature to draw a conclusive connection between the frequency of confirmed germline CNPs and cancer risk based on the data from these 23 subjects.

ACKNOWLEDGMENTS

The study is partially supported by National Institutes of Health grants CA106523 (to J.X.), CA95052 (to J.X.), and Department of Defense grant PC051264 (to J.X.).

REFERENCES

1. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, Chi M, Navin N, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A, Wigler M. 2004. Large-scale copy number polymorphism in the human genome. Science 305:525–528.
2. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C. 2004. Detection of large-scale variation in the human genome. Nat Genet 36:949–951.
3. Tuzun E, Sharp AJ, Bailey JA, Kaul R, Morrison VA, Pertz LM, Haugen E, Hayden H, Albertson D, Pinkel D, Olson MV, Eichler EE. 2005. Fine-scale structural variation of the human genome. Nat Genet 37:727–732.
4. Sharp AJ, Locke DP, McGrath SD, Cheng Z, Bailey JA, Vallente RU, Pertz LM, Clark RA, Schwartz S, Segraves R, Oseroff VV, Albertson DG, Pinkel D, Eichler EE. 2005. Segmental duplications and copy-number variation in the human genome. Am J Hum Genet 77:78–88.
5. Slater HR, Bailey DK, Ren H, Cao M, Bell K, Nasioulas S, Henke R, Choo KH, Kennedy GC. 2005. High-resolution identification of chromosomal abnormalities using oligonucleotide arrays containing 116,204 SNPs. Am J Hum Genet 77:709–726.
6. Xu J, Gillanders EM, Isaacs SD, Chang BL, Wiley KE, Zheng SL, Jones M, Gildea D, Riedesel E, Albertus J, Freas-Lutz D, Markey C, Meyers DA, Walsh PC, Trent JM, Isaacs WB. 2003. Genome-wide scan for prostate cancer susceptibility genes in the Johns Hopkins hereditary prostate cancer families. Prostate 57:320–325.
7. Nannya Y, Sanada M, Nakazaki K, Hosoya N, Wang L, Hangaishi A, Kurokawa M, Chiba S, Bailey DK, Kennedy GC, Ogawa S. 2005. A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays. Cancer Res 65:6071–6079.
8. Lin M, Wei LJ, Sellers WR, Lieberfarb M, Wong WH, Li C. 2004. dChipSNP: Significance curve and clustering of SNP-array-based loss-of-heterozygosity data. Bioinformatics 20:1233–1240.
9. Conrad DF, Andrews TD, Carter NP, Hurles ME, Pritchard JK. 2006. A high-resolution survey of deletion polymorphism in the human genome. Nat Genet 38:75–81.
10. Hinds DA, Kloek AP, Jen M, Chen X, Frazer KA. 2006. Common deletions and SNPs are in linkage disequilibrium in the human genome. Nat Genet 38:82–85.
11. McCarroll SA, Hadnott TN, Perry GH, Sabeti PC, Zody MC, Barrett JC, Dallaire S, Gabriel SB, Lee C, Daly MJ, Altshuler DM, The International HapMap Consortium. 2006. Common deletion polymorphisms in the human genome. Nat Genet 38:86–92.
12. Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST, Bracken MB, Ferris FL, Ott J, Barnstable C, Hoh J. 2005. Complement factor H polymorphism in age-related macular degeneration. Science 308:385–389.

Deletion of a Small Consensus Region at 6q15, Including the *MAP3K7* Gene, Is Significantly Associated with High-Grade Prostate Cancers

Wennuan Liu,[1] Bao-Li Chang,[1] Scott Cramer,[2] Patrick P. Koty,[3] Tao Li,[1] Jishan Sun,[1] Aubrey R. Turner,[1] Chris Von Kap-Herr,[3] Peggy Bobby,[3] Jianyu Rao,[4] S. Lilly Zheng,[1] William B. Isaacs,[5] and Jianfeng Xu[1]

Abstract

Purpose: Chromosome 6q14-21 is commonly deleted in prostate cancers, occurring in ~22% of all tumors and ~40% of metastatic tumors. However, candidate prostate tumor suppressor genes in this region have not been identified, in part due to the large and broad nature of the deleted region implicated in previous studies.

Experimental Design: We first used high-resolution Affymetrix single nucleotide polymorphism arrays to examine DNA from malignant and matched nonmalignant cells from 55 prostate cancer patients. We identified a small consensus region on 6q14-21 and evaluated the deletion status within the region among additional 40 tumors and normal pairs using quantitative PCR and fluorescence *in situ* hybridization. We finally tested the association between the deletion and Gleason score using the Fisher's exact test.

Results: Tumors with small, interstitial deletions at 6q14-21 defined an 817-kb consensus region that is affected in 20 of 21 tumors. The *MAP3K7* gene is one of five genes located in this region. In total, *MAP3K7* was deleted in 32% of 95 tumors. Importantly, deletion of *MAP3K7* was highly associated with higher-grade disease, occurring in 61% of tumors with Gleason score ≥8 compared with only 22% of tumors with Gleason score ≤7. The difference was highly significant ($P = 0.001$).

Conclusion: Our study provides strong evidence for the first time that a small deletion at 6q15, including the *MAP3K7* gene and four other genes, is associated with high-grade prostate cancers. Although the deletion may be a marker for high-grade prostate cancer, additional studies are needed to understand its molecular mechanisms.

Many if not most prostate cancers do not pose a major health threat to their hosts. The molecular factors responsible for variations in the aggressiveness of prostate cancers are poorly defined. Deletion of DNA sequences from chromosome 6q14-21 is one of the most common deletion events in the genome of prostate tumors (reviewed in ref. 1). In a recent study that estimated the frequency of DNA copy number alterations in the prostate cancer genome based on all published comparative genomic hybridization studies of prostate cancers, about one quarter of the 891 prostate cancers had a deletion at 6q14-21 (2). More importantly, the deletion seemed to be more common in metastatic/advanced tumors (40%) than in localized/primary tumors (19%). Despite this overwhelming evidence for frequent 6q deletions, specific prostate tumor suppressor genes have not been identified in this region. One of the major obstacles is the size of the deleted region implicated in previous studies, due at least in part to limited resolution of the detection methods. In our combined analysis of all published comparative genomic hybridization studies, the deleted region at 6q spans ~30 Mb (2). The large number of genes (>170) located within the broad deletion interval poses a significant challenge to effective searches for tumor suppressor genes in the region. Therefore, efforts are needed to define a smaller candidate region. Higher-resolution detection methods, such as representational oligonucleotide microarray analysis and single nucleotide polymorphism (SNP) arrays, may be helpful by identifying smaller deletions and delineating detailed deletion patterns, such as interstitial deletions (3, 4).

In this study, we used high-resolution Affymetrix SNP arrays, fluorescence *in situ* hybridization (FISH), and quantitative PCRs (qPCR) to examine deletion patterns at 6q among 95 prostate tumors. Our goal was to identify a small consensus

Authors' Affiliations: [1]Center for Human Genomics, Departments of [2]Cancer Biology and [3]Pediatrics, Wake Forest University School of Medicine, Winston-Salem, North Carolina; [4]Department of Pathology, University of California at Los Angeles, Los Angeles, California; and [5]Department of Urology, Johns Hopkins Medical Institutions, Baltimore, Maryland

Received 2/6/07; revised 4/10/07; accepted 6/11/07.

Grant support: NIH grants CA106523 (J. Xu), CA95052 (J. Xu), and CA105055 (J. Xu) and Department of Defense grant PC051264 (J. Xu).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

Requests for reprints: Jianfeng Xu, Center for Human Genomics, Wake Forest University School of Medicine, Medical Center Boulevard, Winston-Salem, NC 27157. Phone: 336-713-7500; Fax: 336-713-7566; E-mail: jxu@wfubmc.edu or William B. Isaacs, Department of Urology, Johns Hopkins Medical Institutions, Marburg 115, 600 North Wolfe Street, Baltimore, MD 21287. Phone: 410-955-2518; Fax: 336-502-9336; E-mail: wisaacs@jhmi.edu.

©2007 American Association for Cancer Research.

doi:10.1158/1078-0432.CCR-07-0300

region, evaluate candidate genes within the region using various molecular methods, and examine the possible correlation between gene deletion and tumor aggressiveness.

Materials and Methods

Subjects. All tumors analyzed in the study were from independent prostate cancer patients undergoing radical prostatectomy for treatment of clinically localized disease at Johns Hopkins Hospital. We selected a subset of subjects from whom genomic DNA of sufficient quantity (>5 μg) and purity (>70% cancer cells for cancer specimens, no detectable cancer cells for normal samples) could be obtained by dissection of matched nonmalignant (hereafter referred to as "normal") and cancer containing areas of prostate tissue as determined by histologic evaluation of H&E-stained frozen sections of snap-frozen radical prostatectomy specimens. Genomic DNA was isolated from trimmed frozen tissues as described previously (5).

Detection of 6q deletions using Affymetrix SNP arrays. We first used Affymetrix SNP array panels to detect DNA copy number alterations in the genome among 55 prostate cancers, and for this study, we focused entirely on 6q deletions. We used the 100K SNP array for the first 22 subjects (4) and then used the 500K SNP array (Affymetrix, Inc.) for the remaining 41 subjects. Eight of the subjects were analyzed using both 100K and 500K SNP arrays. All of the reagents used for the assay were obtained from the manufacturers recommended by Affymetrix. We digested, ligated, amplified, fragmented, and labeled the samples and hybridized, washed, and stained the SNP mapping arrays according to the manufacturer's instructions. DNA copy number was calculated based on allele intensity using two different software packages: Copy Number Analyzer for Affymetrix GeneChip (CNAG2.0; ref. 6) and dChip analyzer (7). Allele-specific analysis was also done to estimate DNA copy number for each chromosome pair using CNAG2.0. The results from all three types of analyses using CNAG2.0 and dChip were consistent, and we show the output from CNAG2.0 in Fig. 1. The physical positions of detected deletions were determined based on the Human hg17 Assembly (National Center for Biotechnology Information Build 35).

Detection of DNA deletion at MAP3K7 using qPCR. qPCR analysis was used to evaluate the deletion status of the *MAP3K7* gene in the 55 tumors described above and among 40 additional prostate tumors that were not evaluated by SNP arrays. The qPCR analysis was done using the ABI Prism 7000 Sequence Detection System as described in detail elsewhere (4). A primer set located in the last exon of *MAP3K7* was used to amplify the test amplicon (forward primer: 5′-AACGGTCCCAGAGAATCATGAAGTGC-3′; reverse primer: 5′-GAGGTCATCAGAAGTCAGCAGCAGAA-3′), and a primer set located around the junction of intron 2 and exon 3 of *GAPDH* was used to amplify the control amplicon (forward primer: 5′-TCCTCATGCCTTCTTGCCTCTTGT-3′; reverse primer: 5′-AGGCGCCCAATACGACCAAATCTA-3′).

Statistical analysis. The difference in the frequencies of DNA deletion at *MAP3K7* between lower-grade (Gleason scores ≤7) and higher-grade tumors (Gleason scores ≥8) was tested using Fisher's exact test.

Confirmation of DNA deletion at MAP3K7 using FISH. A subset of identified *MAP3K7* deleted tumors was analyzed using FISH to confirm their deletion status. We obtained the PAC clone RP1-154G14 (~100 kb at 6q15-16.3) for the FISH analysis from the Children's Hospital Oakland Research Institute.[6] The only known gene contained in this clone is *MAP3K7*. We grew the clone, isolated the DNA, and checked the identity of the insert in the clone as recommended. The hybridization mixture contained 200 ng of PAC RP1-154G14 DNA, which was labeled by nick translation with SpectrumOrange dUTP following the manufacturer's protocol (Vysis, Inc.). In addition, the hybridization mixture contained two additional probes: the centromeric probe CEP6 labeled with SpectrumGreen (Vysis) and the LSI MYB probe labeled with SpectrumAqua (Vysis), which bracket the PAC RP1-154G14 clone (Fig. 2B). Paraffin-embedded normal prostate and prostate cancer sections (5 μm) were pretreated, hybridized, and washed following the manufacturer's protocol (Vysis). FISH analysis was done using fluorescent microscopy with the appropriate filters to visualize the three probes. Slides were blinded, and then for each sample, a total of 200 interphase cells was analyzed independently by two individuals.

MAP3K7 protein expression using immunohistochemistry. For immunohistochemical staining of proteins, sections were deparaffinized by successive incubation in xylene, 100% ethanol, and 90% ethanol following standard procedures. The endogenous peroxidase activity was blocked by incubation for 20 min at room temperature in 0.5% H_2O_2 in water. Sections were washed thrice in PBS. Retrieval of antigens was done by incubating the sections in antigen retrieval solution (Sigma) for 1 h in a 95°C water bath. After allowing the sections to cool, samples were washed in PBS and blocked with 3% bovine serum albumin in PBS for 30 min at room temperature. After blocking, sections were incubated with anti-transforming growth factor-β–activated kinase 1 (Tak1) primary antibody (1:50 dilution; Abcam, Inc.) for 1 h at room temperature and washed. Sections were then incubated with the goat anti-rabbit secondary antibody with a horseradish peroxidase conjugate (Jackson ImmunoResearch Laboratories, Inc.), washed, and then incubated with diaminobenzidine for approximately 2 to 5 min. Following counterstain with hematoxylin (Sigma) and clearing of the sections through ethanol and xylene, coverslips were mounted using Permount medium.

Results and Discussion

The use of high-resolution SNP arrays and allele-specific analyses improved our ability to detect somatic DNA deletions in prostate tumors. Detectable deletions at 6q14-21 were observed in 21 of the 55 prostate cancers (38%) examined using the SNP arrays in this study (Fig. 1). This frequency was considerably higher than the estimate of 25% obtained from a combined analysis of 891 prostate tumors described in published comparative genomic hybridization studies (2). One reason for the higher estimate of deletion in our study is the use of a higher-resolution detection method. For example, we were able to detect a small deletion of ~800 kb at 6q15 in tumor G7-042 using the SNP arrays (Fig. 1). Another potential factor for the higher estimate of deletion was the greater proportion of tumors with high-grade disease (38% with Gleason score ≥8) in our study. We found that the frequency of 6q deletion was significantly higher in tumors with Gleason score ≥8 (12 of 17 tumors, 71%) than that of Gleason scores ≤7 (9 of 38, 24%; $P = 0.002$). The association between 6q deletion and tumor grade was striking; in fact, it was the strongest among all the association tests between any common recurrent DNA copy number changes (defined as ≥10%) identified in our studies and tumor grade.

The high-resolution SNP arrays and the ability to analyze copy number changes for each chromosome using allele-specific analysis also improved our ability to dissect detailed deletion patterns. We detected three small-size homozygous deletions among these 21 tumors with 6q deletions (Fig. 1). One was ~838 kb (98,588-99,426 kb) in tumor G7-019 and contained the *POU3F2* gene, one was ~265 kb (117,394-117,659 kb) in tumor G7-048 and contained no known gene, and the other was ~1.1 Mb (84,868-85,963 kb) in

[6] BACPACorders@chori.org

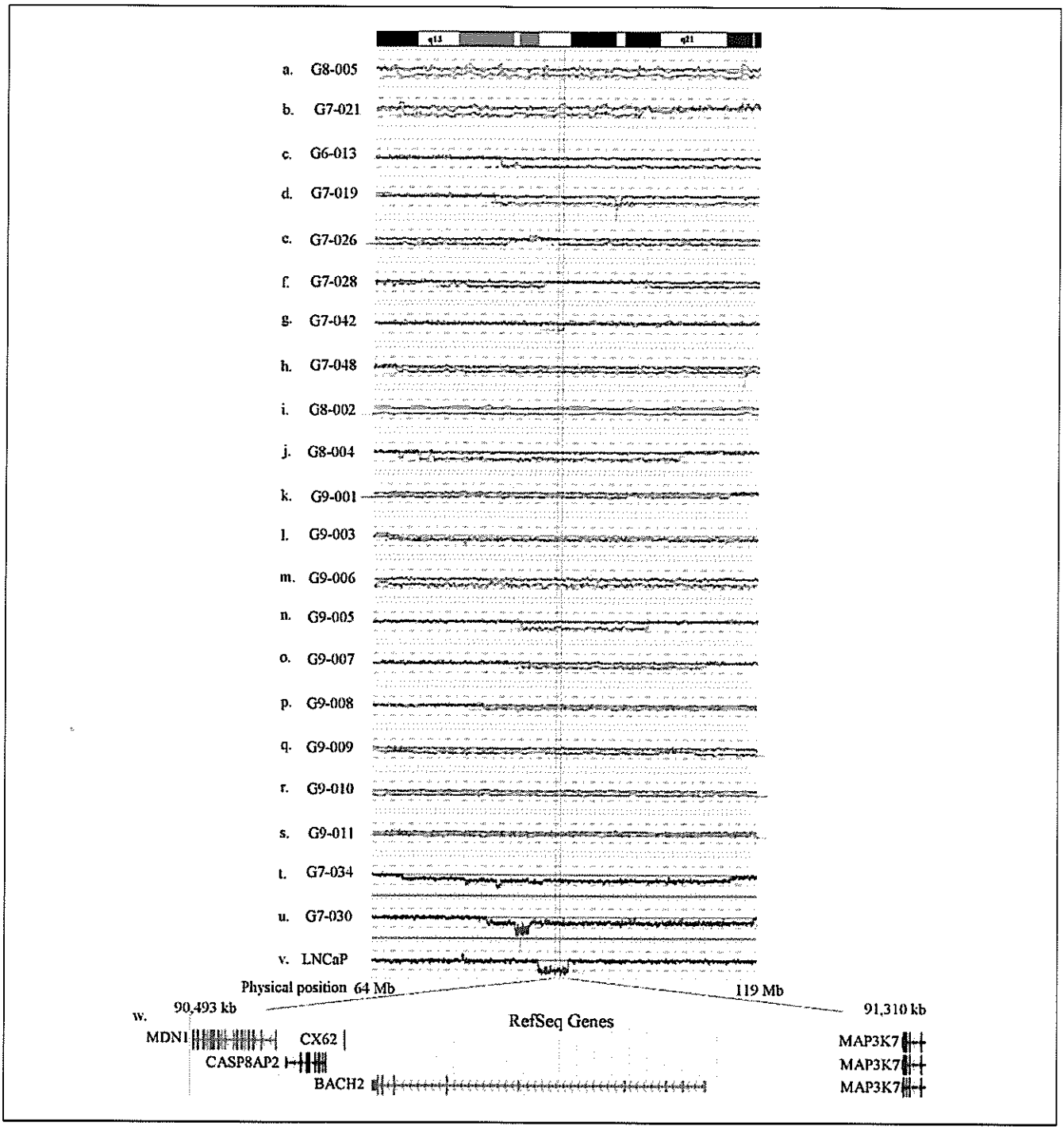

Fig. 1. Examples of genomic deletions on chromosome 6q identified in tumor genomes using Affymetrix 100K (*a* and *b*) and 500K (*c-v*) SNP mapping arrays, and CNAG2.0 with genomic smoothing of 10 SNPs. a to s, allele-specific analysis using matched normal DNA as reference. t to v, non – allele-specific analysis using automatically selected references by CNAG2.0 from our database containing 40 normal samples. d, from Nsp array only. i, from Sty array only. Hemizygous deletions outlined by green (allele-specific analysis) or blue (non – allele-specific analysis) horizontal curves below the baseline (*solid black line*) are marked by green arrows. Homozygous deletions outlined by red and green or blue horizontal curves below the baseline (*solid black line*) are marked by red arrows. Vertical dotted lines, minimum overlapping deleted region. w, genes in the minimum overlapping deleted region based on the Human hg17 Assembly (University of California at Santa Cruz, National Center for Biotechnology Information Build 35).

tumor G7-030 and contained the *Q9Y2L2* gene. These homozygous deletions, however, did not overlap. Although the significance of these homozygous deletions in prostate cancer development is unclear, their nonoverlapping nature may be more consistent with generalized genomic instability than a specific selection for loss of these particular genomic regions.

In addition, we found two additional tumors (tumors G7-026 and G7-028) with interstitial deletions at 6q (Fig. 1). The distal deleted region of tumor G7-026, between 90,451 and

97,338 kb, provided information that was critical in defining a minimal overlapping region (see below). When examining the deleted regions shared by each of these 21 tumors, we found one small region, between 90,493 and 91,310 kb, which was shared by all but one tumor. Interestingly, this shared deleted region fell within the 4.3 Mb 6q deleted region observed in LNCaP and seemed to be consistent with the minimum loss of heterozygosity regions reported by Cooney et al. (8), Srikantan et al. (9), Hyytinen et al. (10), and Konishi et al. (11). Only five genes are located within this region of 817 kb, including *MDN1*, *CASP8AP2*, *CX62*, *BACH2*, and *MAP3K7*. To further evaluate the significance of these five genes, we examined their differential expression patterns in the ONCOMINE gene expression database.[7] The expression of the *MAP3K7* gene was significantly lower in the tumor cells in comparison with the normal tissues of the prostate ($P = 0.0001$). In contrast, the expressions of the other four genes in prostate tumors were not significantly different from normal. The *MAP3K7* gene encodes Tak1, a member of the mitogen-activated protein kinase kinase kinase family and an activator of c-Jun NH_2-terminal kinase and p38 mitogen-activated protein kinase pathways (12). Because of the decreased expression of *MAP3K7* in prostate cancer and its potential tumor suppressor role, we targeted the *MAP3K7* gene in the remaining analyses.

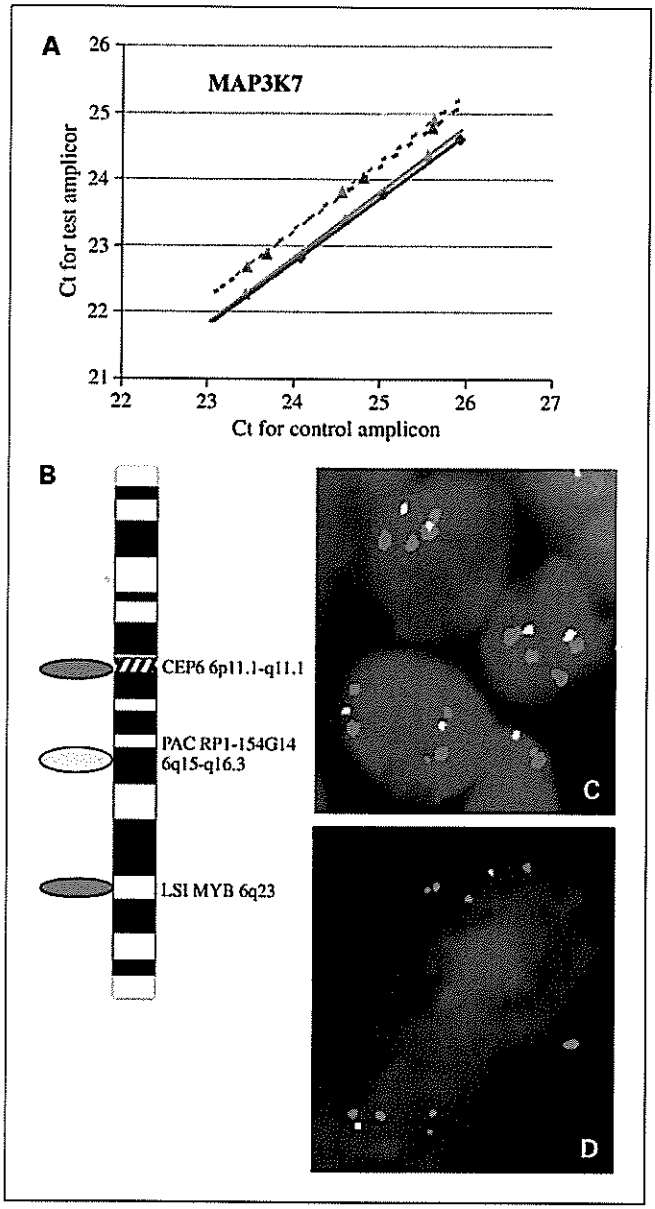

Fig. 2. Confirmation of *MAP3K7* deletions in various tumors using qPCR and FISH. *A*, examples of validating *MAP3K7* deletions by real-time qPCR. Solid lines, controls; dotted lines, tumors; blue, subject G8-005; red, subject G9-003. *B*, ideogram of chromosome 6 showing the physical map location of the FISH probes. Green, CEP6 probe; orange, PAC RP1-154G14 clone; aqua, LSI MYB. *C*, normal prostate control with a normal signal pattern (2G2O2A). *D*, tumor G8-005 with a signal pattern indicating a hemizygous deletion of the PAC RP1-154G14 clone (2G1O2A).

To confirm that the *MAP3K7* gene was implicated in these 6q14-21 deleted prostate tumors, we did a qPCR analysis using a probe located in the last exon of the *MAP3K7* gene. As expected, qPCR analysis detected a hemizygous deletion in the 20 tumors that shared the 817-kb minimal overlapped deleted region (Fig. 2A) and also confirmed that the remaining tumor (tumor G7-028) does not have a deletion at this region. To further confirm the SNP array and qPCR results, we did FISH analysis in a subset of samples (Fig. 2B-D). The RP1-154G14 (~100 kb and includes the *MAP3K7* gene), CEP6, and LSI MYB probes were hybridized to tumors G8-005, G9-003, and G7-019, and the probe signal was analyzed in 200 interphase cells for each sample (Fig. 2C and D). This analysis revealed a pattern consistent with a hemizygous interstitial deletion of the *MAP3K7* region of chromosome 6q15-q16.3 (2G1O2A signal pattern, as shown in Fig. 2D) in tumor G8-005 (54%), tumor G9-003 (82%), and tumor G7-019 (55%), whereas the normal prostate control revealed a normal signal pattern (2G2O2A; Fig. 2C) in 94% of the cells. This analysis confirmed the SNP array and qPCR results, which also identified a hemizygous interstitial deletion of the *MAP3K7* region of chromosome 6q15-q16.3.

To obtain a better estimate of *MAP3K7* deletion frequency in prostate tumors and assess their association with Gleason scores, we used the same qPCR assay to examine the deletion status in 40 additional tumors that have not been analyzed by SNP arrays. The vast majority of these tumors had a Gleason score of 6 or 7. We observed the *MAP3K7* gene deletion in nine of these tumors. In total, the *MAP3K7* gene deletion was detected in 30 of the 95 (32%) tumors in our study using either SNP array or qPCR method. Importantly, the deletion was considerably more common in higher-grade tumors (14 of 23 tumors with Gleason score ≥ 8, 61%) than in lower-grade tumors (16 of 72 tumors with Gleason score ≤ 7, 22%; Table 1). The difference was highly significant ($P = 0.001$). Most strikingly, the frequency of the *MAP3K7* deletion was highest in Gleason 9 tumors, occurring in 9 of the 12 tumors (75%).

As a test of our genetic data, we immunostained representative prostate tissue samples with anti-Tak (alias *MAP3K7*) antibody. Figure 3 shows immunostaining of benign versus high-grade tumor (Gleason 9) in a sample (tumor G9-001) with confirmed hemizygous deletion of *MAP3K7*. Benign tissue shows staining in the cytoplasm and at the plasma membrane

[7] http://www.oncomine.org

Table 1. *MAP3K7* deletion in prostate tumors

Group	Gleason sum	No. tumors examined	*MAP3K7* deletion	
			No. tumors	% tumors
Lower-grade tumor	6	19	4	21
	3 + 4	36	9	25
	4 + 3	17	3	18
	Subtotal	72	16	22
Higher-grade tumor	8	11	5	45
	9	12	9	75
	Subtotal	23	14	61
All tumors		95	30	32

of prostatic epithelial cells. The intensity of staining ranges from moderate (data not shown) to intense (Fig. 3A). There are also moderate to strongly staining stromal cells present. In the sample depicted in Fig. 3, the immunoreactivity present in high-grade tumor is greatly diminished (Fig. 3B), consistent with the hemizygous genotype of this sample. Also note the strong immunoreactivity of stromal cells in the panel of the high-grade specimen. Two additional samples with verified hemizygous genotype showed similar Tak1 immunoreactivity (data not shown). These results are consistent with the findings that *MAP3K7* RNA expression level is greatly reduced in prostate cancers in comparison with the level in normal prostate tissues.[7]

Our finding that a small consensus deleted region in prostate cancers, including the *MAP3K7* gene, is strongly associated with Gleason score is significant. The *MAP3K7* gene encodes Tak1. Tak1 is a member of the mitogen-activated protein kinase kinase kinase family that was originally identified as a key regulator of mitogen-activated protein kinase activation in transforming growth factor-β–induced signaling pathways (12). Mutations and alterations in several layers of the transforming growth factor-β signaling cascade have been identified, including ligand, receptors, and intracellular signaling events (13). Before our genetic study, Tak1 has not been implicated in this process. Our demonstration that the *MAP3K7* locus is deleted in prostate tumor specimens adds another step in transforming growth factor-β signaling that is abrogated in prostate tumorigenesis and further strengthens the role of this pathway in prostate cancer development.

Although Gleason score is perhaps the most reliable predictor for prostate cancer behavior, it is far from ideal. Many low- to intermediate-grade cancers may be metastatic, whereas a few of the high-grade tumors still have an indolent course. Thus, molecular markers that could further define the disease prognosis are greatly needed. Markers that can be used, either alone or in conjunction with Gleason score, to more precisely identify prostate cancers capable of progression to disseminated disease would be extremely useful in determining which patients to treat and how aggressively. It is interesting to note that the association between *MAP3K7* deletion and Gleason grade is stronger than that of between p53 and Gleason grade in our study. Deletion of the p53 region at 17p13 occurs in 52.94% of tumors with Gleason score ≥8 and in 16.79% of tumors with Gleason score ≤7 ($P = 0.008$). In comparison, the deletion at *MAP3K7* region occurs in 61% of tumors with Gleason score ≥8 and only in 22% of tumors with Gleason score ≤7 ($P = 0.001$). Further work in larger patient populations with follow-up information will be needed to assess the potential of *MAR3K7* deletion as a prognostic marker.

Although we have shown the significance of the *MAP3K7* deletion in prostate cancer, these results are subject to several caveats. First, our data only provided a strong statistical evidence for association between a consensus deleted region, including the *MAP3K7* gene, and tumor grade; we did not assess the causal relationship between *MAP3K7* gene and aggressive prostate cancer. Other *in vitro*, *in vivo*, and animal studies are needed to understand its molecular mechanisms. Second, the current study targeted only the *MAP3K7* gene, and therefore, the roles of four other genes at this minimum overlap region in prostate cancer development remain to be

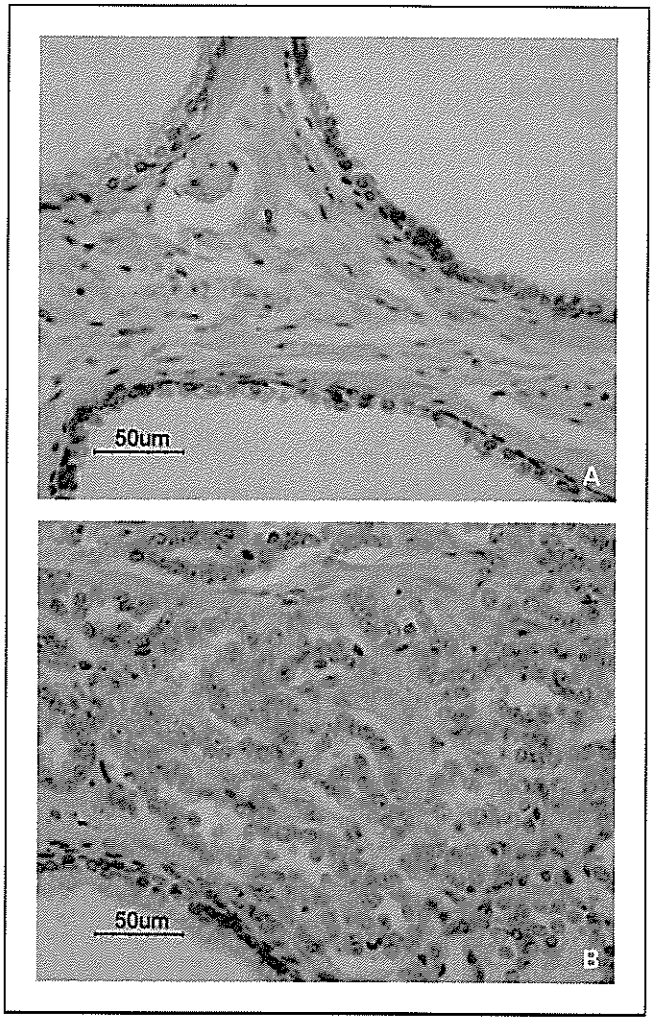

Fig. 3. Tak1 immunoreactivity is decreased in high-grade prostate tumors. Sections (5 μm) were stained with rabbit anti-Tak1 antibody followed by labeling protocols using goat anti-rabbit horseradish peroxidase secondary as described in Materials and Methods. Controls that lacked primary antibody were clear (data not shown). The images shown were from different areas of the same section of a specimen verified by 100K SNP and q-PCR to be hemizygous for the *MAP3K7* locus. *A*, benign region. *B*, Gleason 5 tumor cells. Note benign gland in lower-right hand corner of (*B*) showing positive Tak1 immunoreactivity.

further investigated. Finally, our results differ from a previous study that found no correlation between loss of heterozygosity at 6q and *MAP3K7* transcript expression (11). The differences in the deletion detection methods [loss of heterozygosity using microsatellite markers in the study of Konishi et al. (11) versus SNP arrays and qPCR in our study] and sample size [21 subjects in the study of Konishi et al. (11) versus 95 patients in our study] may contribute to the contradictory findings.

In summary, our study provides evidence that deletion of an 817-kb region of 6q15, including the *MAP3K7* gene, is one of the most consistent genetic alterations occurring in the genome of high-grade prostate cancers. Although *MAP3K7* represents an important candidate prostate tumor suppressor gene affected by this genomic alteration, further studies will be necessary to establish a causal relationship between this gene and/or the other genes in this interval and prostate cancer initiation and/or progression.

References

1. Dong JT. Chromosomal deletions and tumor suppressor genes in prostate cancer. Cancer Metastasis Rev 2001;20:173–93.
2. Sun J, Liu W, Adams TA, et al. DNA copy number alterations in prostate cancers: a combined analysis of published CGH studies. Prostate 2007;67:692–700.
3. Lucito R, Healy J, Alexander J, et al. Representational oligonucleotide microarray analysis: a high-resolution method to detect genome copy number variation. Genome Res 2003;13:2291–305.
4. Liu W, Chang B, Sauvageot J, et al. Comprehensive assessment of DNA copy number alterations in human prostate cancers using Affymetrix 100K SNP mapping array. Genes Chromosomes Cancer 2006; 45:1018–32.
5. Bova GS, MacGrogan D, Levy A, Pin SS, Bookstein R, Isaacs WB. Physical mapping of chromosome 8p22 markers and their homozygous deletion in a metastatic prostate cancer. Genomics 1996;35:46–54.
6. Nannya Y, Sanada M, Nakazaki K, et al. A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays. Cancer Res 2005;65:6071–9.
7. Lin M, Wei LJ, Sellers WR, Lieberfarb M, Wong WH, Li C. dChipSNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data. Bioinformatics 2004;20:1233–40.
8. Cooney KA, Wetzel JC, Consolino CM, Wojno KJ. Identification and characterization of proximal 6q deletions in prostate cancer. Cancer Res 1996;56: 4150–3.
9. Srikantan V, Sesterhenn IA, Davis L, et al. Allelic loss on chromosome 6Q in primary prostate cancer. Int J Cancer 1999;84:331–5.
10. Hyytinen ER, Saadut R, Chen C, et al. Defining the region(s) of deletion at 6q16-q22 in human prostate cancer. Genes Chromosomes Cancer 2002;34:306–12.
11. Konishi N, Nakamura M, Kishi M, et al. Genetic mapping of allelic loss on chromosome 6q within heterogeneous prostate carcinoma. Cancer Sci 2003;94: 764–8.
12. Yamaguchi K, Shirakabe K, Shibuya H, et al. Identification of a member of the MAPKKK family as a potential mediator of TGF-β signal transduction. Science 1995;270:2008–11.
13. Bierie B, Moses HL. TGF-β and cancer. Cytokine Growth Factor Rev 2006;17:29–40.

GENES, CHROMOSOMES & CANCER 46:972–980 (2007)

Multiple Genomic Alterations on 21q22 Predict Various *TMPRSS2/ERG* Fusion Transcripts in Human Prostate Cancers

Wennuan Liu,[1] Charles M. Ewing,[2] Bao-Li Chang,[1] Tao Li,[1] Jishan Sun,[1] Aubrey R. Turner,[1] Latchezar Dimitrov,[1] Yi Zhu,[1] Jielin Sun,[1] Jin Woo Kim,[1] S. Lilly Zheng,[1] William B. Isaacs,[2]* and Jianfeng Xu[1]*

[1]Center for Human Genomics, Wake Forest University School of Medicine, Winston-Salem, NC
[2]Johns Hopkins Medical Institutions, Baltimore, MD

A number of *TMPRSS2/ERG* fusion transcripts have been reported since the discovery that recurrent genomic rearrangements result in the fusion of *TMPRSS2* and ETS family member genes. In this article we present evidence demonstrating that multiple genomic alterations contribute to the formation of various *TMPRSS2/ERG* transcripts. Using allele-specific analysis of the data generated from the GeneChip 500K SNP array we observed both hemizygous and homozygous deletions occurring at different locations between and within *TMPRSS2* and *ERG* in prostate cancers. The 500K SNP array enabled us to fine map the start and end of each deletion to specific introns of these two genes, and to predict a variety of fusion transcripts, including a new form which was confirmed by sequence analysis of the fusion transcripts in various tumors. We also inferred that translocation is an additional mechanism of fusion for these two genes in some tumors, based on largely diploid genomic DNA between *TMPRSS* and *ERG*, and different fusion transcripts produced in these tumors. Using a bioinformatics approach, we then uncovered the consensus sequences in the regions harboring the breakpoints of the deletions. These consensus sequences were homologous to the human Alu-Sq and Alu-Sp subfamily consensus sequences, with more than 80% homology. The presence/absence of Alu family consensus sequence in the introns of *TMPRSS2* and *ERG* correlates with the presence/absence of fusion transcripts of theses two genes, indicating that these consensus sequences may contribute to genomic deletions and the fusion of *TMPRSS2* and *ERG* in prostate cancer. © 2007 Wiley-Liss, Inc.

INTRODUCTION

The recent discovery of recurrent chromosomal rearrangements resulting in the fusion of androgen regulated *TMPRSS2* to transcripts of ETS family member genes in the majority of prostate cancers has not only firmly established the existence of genomic rearrangements occurring in epithelial malignancies, but could also shed light on the progression of this clinically heterogeneous disease. In a study analyzing paired benign and malignant prostate epithelial cells from 114 prostate cancer patients using expression microarray, Petrovics et al. (2005) identified the ETS-related gene *ERG* as the most frequently overexpressed proto-oncogene in the transcriptome of malignant prostate epithelial cells. In many cases, the overexpression of the ETS family of genes, including *ERG*, *ETV1*, and *ETV4* is apparently caused by the fusion of *TMPRSS2*, a gene upregulated by androgens in prostate cancer cells, to ETS transcription factor genes (Tomlins et al., 2005; Cerveira et al., 2006; Perner et al., 2006; Wang et al., 2006; Yoshimoto et al., 2006; Clark et al., 2007). Although no clinical characteristics related to the fusion were observed by Yoshimoto et al. (2006), both Perner et al. (2006) and Wang et al. (2006) have reported a positive association between *TMPRSS2/ERG* fusion and clinical outcomes of prostate cancer progression. While there is marked heterogeneity in the expressed forms of *TMPRSS2/ERG* transcripts in different tumors (Clark et al., 2007), Wang et al. (2006) found that the expression of various fusion transcripts was correlated with clinical outcome of prostate cancers. It is possible that these markedly different isoforms may, at least in part, reflect differences in genomic architecture resulting from translocations and genomic fusion at different positions.

Supported by: National Cancer Institute, Grant numbers: CA106523, CA95052; Department of Defense, Grant number: PC051264.

*Correspondence to: William B. Isaacs, Johns Hopkins Hospital, Marburg 115, 600 N. Wolfe Street, Baltimore, MD 21287, USA. E-mail: wisaacs@jhmi.edu or Jianfeng Xu, Center for Human Genomics, Medical Center Blvd, Winston-Salem, NC 27157, USA. E-mail: jxu@wfubmc.edu

Received 27 March 2007; Accepted 29 June 2007
DOI 10.1002/gcc.20482
Published online 25 July 2007 in Wiley InterScience (www.interscience.wiley.com).

© 2007 Wiley-Liss, Inc.

WILEY InterScience® DISCOVER SOMETHING GREAT

While the causes of these translocations on chromosome 21 in prostate tumors remain to be elucidated, we and others have demonstrated the existence of genomic deletions between *TMPRSS2* and *ERG* using 100K SNP arrays and FISH (Liu et al., 2006; Perner et al., 2006; Yoshimoto et al., 2006). Although the DNA breakpoints were not identified in these primary tumors due to the limited resolution of the techniques used in these studies, Hermans et al. (2006) have mapped the breakpoint in intron 2 of *ERG* and introns 1 and 2 of *TMPRSS2* in xenograft DNAs using long-range PCR. In addition, the effects of the loss of one copy of the genes located between *TMPRSS2* and *ERG* remain to be fully evaluated. However, a knocking out *HMGN1*, one of the 14 known genes located between *TMPRSS2* and *ERG*, in a mouse model has been shown to result in increased levels of N-cadherin expression (Rubinstein et al., 2005), which is a characteristic of high-grade prostate cancer (Tomita et al., 2000).

To further characterize the genomic alterations involved in the fusion of these two genes, we now report our findings from 41 pairs of primary prostate tumors and matched nonmalignant tissues, evaluated using the GeneChip mapping 500K SNP array and allele-specific analysis. Our results demonstrate that multiple genomic alterations/translocations, at least in part, contribute to the diversity of the fusion transcripts found in various prostate cancers.

MATERIAL AND METHODS

Study Subjects

All subjects for the somatic DNA deletion analysis of prostate tumors were prostate cancer patients undergoing radical prostatectomy (RP) for treatment of clinically localized disease at Johns Hopkins Hospital. All subjects participated in an IRB approved protocol. We selected 41 subjects from whom genomic DNA of sufficient amount (>5 μg) and purity (>70% cancer cells for cancer specimens, no detectable cancer cells for normal samples) could be obtained by dissection of matched nonmalignant (hereafter referred to as "normal") and cancer containing areas of prostate tissue as determined by histological evaluation of H&E-stained frozen sections of snap frozen RP specimens. Genomic DNA was isolated from trimmed frozen tissues as previously described (Bova et al., 1993).

GeneChip Mapping 500K SNP Assay

The GeneChip Mapping 500K set is comprised of Nsp (~262,000 SNPs) and Sty (~238,000 SNPs) arrays. The median physical distance between SNPs is ~2.5 kb and the average distance between SNPs is ~5.8 kb. The 500K SNP arrays were purchased from Affymetrix, Santa Clara, CA. All of the reagents used for the assay were obtained from manufacturers recommended by Affymetrix. The 500K SNP mapping arrays were labeled, hybridized, washed, and stained according to the manufacturer's instructions. Briefly, 250 ng of genomic DNA was digested with either *NspI* or *StyI* and then ligated to adapters. A generic primer that recognizes the adapter sequence was used to amplify adapter ligated DNA fragments with PCR conditions optimized to preferentially amplify fragments in the 200 to 1,100 bp size range in a GeneAmp PCR System 9700 (Applied Biosystems, Foster City, CA). After purification with a Clontech DNA amplification clean up kit, a total of 90 μg of PCR product was fragmented and a sample of the fragmented product was visualized on a 4% TBE agarose gel to confirm that the average size was smaller than 180 bp. The fragmented DNA was then labeled with biotin and hybridized to the arrays for 18 hr. The arrays were washed and stained using an Affymetrix Fluidics Station 450 and scanned using a GeneChip Scanner 3000 7G (Affymetrix). The allele intensity of each SNP was measured using the GeneChip Genotyping analysis software (GTYPE). DNA copy number was calculated based on allele intensity using two different software packages; Copy Number Analyzer for Affymetrix GeneChip (CNAG2.0; Nannya et al., 2005), and dChip analyzer (dChip; Lin et al., 2004). Allele-specific analysis was also performed to estimate DNA copy number for each of heterozygous alleles using CNAG2.0. The physical positions of the detected deletions were determined based on the Human hg17 Assembly (NCBI Build 35).

Quantitative PCR (qPCR)

qPCR was performed using the ABI Prism 7500 Sequence Detection System. Primers were designed using Primer Express 1.5 software from Applied Biosystems. Amplicons were designed against the locus that was putatively homozygous deleted in the tumor and a control locus of known normal DNA CN. The sequences of the primers for *TMPRSS2* and *ERG* are available upon request. PCR kinetics at the control locus were used to control for sample-to-sample differences in genomic DNA purity and concentration. Three concentrations of each genomic DNA sample (20, 10, and 5 ng) were assayed in duplicate, using each pair of real-time PCR primers. PCRs were prepared as fol-

lows: in 20 μl, we combined 2 μl of genomic DNA, 0.05 μM of each primer, and SYBR-Green PCR Master Mix from Applied Biosystems. PCR reactions were performed as follows: 95°C for 10 min, followed by 40 cycles at 95°C for 20 sec, and 60°C for 1 min. An additional cycle of 95°C for 15 sec, 60°C for 20 sec, and 95°C for 15 sec was run at the end to measure the dissociation curve for quality control. We used Sequence Detection Software (SDS) for PCR baseline subtraction and then exported the threshold cycle number (Ct) data for analysis. Ct values of the control and test amplicons for the three dilutions of each DNA sample were plotted against each other, and the offset between the two samples, along the control-amplicon axis and test-amplicon axis were measured to determine the status of genomic deletions.

Analysis of Fusion Transcripts

Total RNA was isolated from 10-μm sections of prostate tumors using TRIZOL reagent (Invitrogen, Carlsbad, CA) according to the manufacturer's instructions. We synthesized cDNA from the RNA using a cDNA Synthesis System from Roche Applied Science (Indianapolis, IN) according to the manufacturer's protocol. RT-PCR was performed in a 50 μl reaction volume for 1 cycle at 95°C for 2 min, then 38 cycles at 95°C for 30 sec, 58°C for 30 sec and 72°C for 1 min, and ending with 10 min at 72°C. We used ERG-V_F (cacggtaatgcatgctagaa) and ERG-V_R (ggttgagcagctttcg actg) for expression analysis of ERG variant 1 (NM_182918) and ERG variant 2 (NM_004449). We used the primers t-e_A F (CAGGAGGCG GAGGCGGA) and t-e_A R (GGCGTTGTAGC TGGGGGTGAG), t-e_C F1 (CAGCAAGATGG CTTTGAACTC) and t-e_C R1 (GGATCTGC TGGCACGATAAC), respectively, for analyses of the fusion transcripts in various tumors.

The sequencing reactions were performed using dye-terminator chemistry (BigDye, ABI, Foster City, CA). The reaction was performed in a total volume of 5 μl with about 10–30 ng of purified PCR products, 0.5 μl of BigDye and 0.16 μM of forward or reverse PCR primers (t-e_A F, t-e_A R, t-e_C F: GATAACAGCAAGATGGCTTTG 3′, t-e_C R: GATCTGCTGGCACGATAACTC). The sequencing cycling conditions were as follows: 95°C for 30 sec; followed by 22 cycles of 95°C for 30 sec, 50°C for 10 sec and 60°C for 4 min. After sequencing cycling the samples were precipitated using 63 ± 5% ethanol. Then 10 μl of Hi-di formamide (ABI, Foster City, CA) was added before samples were loaded onto an ABI 3730 DNA Analyzer. Sequencher™ software version 4.1.4 (Gene Codes Corporation) was used for sequence analysis.

RESULTS

Genomic Deletions Between TMPRSS2 and ERG

In a previous study using the 100K SNP array, we noticed that the centromeric boundaries of deletions between *TMPRSS2* and *ERG* varied among different prostate tumors, while it was impossible to compare the boundaries on the telomeric side because of poor resolution. In this study, we increased the number of study subjects by including 33 normal-tumor pairs of new samples, and we used the 500K mapping array to refine the boundaries of various deletions in this region. In order to elucidate the detailed genomic structures involved in the deletions, we also included eight of the tumors that had been previously analyzed using the 100K SNP array. These include G6-002T, G6-003T, G7-002T, G7-009T, G7-017T, G7-032T, G8-001T, and G8-002T. While we did not present the deletions between *TMPRSS2* and *ERG* (T-E deletion) in G6-002T, G7-002T, G7-032T, and G8-002T in our previous publication, we reported T-E deletions in the rest of the tumors (Liu et al., 2006). Using the 500K mapping array in this study, we not only refined T-E deletions in G6-002T, G7-002T, G7-032T, and G8-002T but also identified 8 more subjects harboring T-E deletions among the 33 pairs of new samples (Fig. 1).

Homozygous and Hemizygous Deletions Revealed by Allele-Specific Analysis

To characterize the deletions involving *TMPRSS2*, *ERG*, and the genes in between, we employed allele-specific analysis as described by Nannya et al. (2005). The algorithm of allele-specific analysis takes advantage of genotype information and allele-specific intensities from paired samples to estimate DNA copy numbers for each heterozygous SNP. Allele-specific analysis can also minimize the effect of normal DNA contamination from nonmalignant cells in the tumors on the identification of DNA copy number changes, because the other allele can be used as an internal control for hemizygous deletion or amplification.

Using allele-specific analysis, we analyzed a total of 41 pairs of samples from primary prostate tumors and matched nonmalignant tissues as described above with 16 tumors harboring T-E deletion (Supplementary Table 1, http://www1.wfubmc.edu/Genomics/Publications+and+Data). Of the genomic losses between *TMPRSS2* and *ERG*, 13 tumors apparently harbored hemizygous deletions as shown

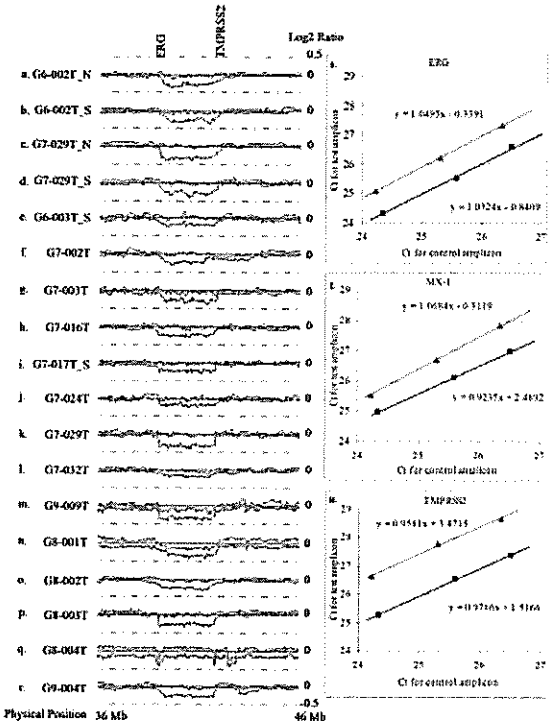

Figure 1. DNA copy number changes on chromosome 21 revealed by allele-specific analysis of the genotype and allele intensity data generated from the GeneChip 500K SNP assay using CNAG2.0 in prostate tumors. (a–d) log2 ratio analysis with 10-SNP genomic smoothing shows various genomic deletions between *ERG* and *TMPRSS2*, as detected by two independent Nsp and Sty arrays. The consistency of the results from these two independent assays confirmed the deletion in this region. (e–r) allele-specific analysis of the 500K (Nsp and Sty) SNP data illustrates genomic deletions between *ERG* and *TMPRSS2* in different prostate tumors. G6-002T (k), G6-003T (e), and G7-017T (i) were analyzed only in the Sty SNP array. Red and green curves represent different alleles. (s–u) validating deletions of *ERG* (s), *MX1* (t), and *TMPRSS2* (u) in Tumor G9-009T by real-time quantitative PCR. Ct values of the control (*GAPDH*, *x*-axis) and test (*y*-axis) amplicons for the three dilutions of each DNA sample were plotted against each other, and the offsets between best-fit lines for the samples along the test-amplicon axis at 25 Ct of the control-amplicon axis were measured. The offsets in the Ct values (ΔCt) between tumor DNAs with putative *ERG*, *MX1*, or *TMPRSS2* deletions and paired normal DNAs without deletion in these regions are used to infer DNA copy number change. Tumor DNA is defined as having a hemizygous deletion when the ΔCt is less than 0.68 and as having a homozygous deletion when the ΔCt is more than 0.68, assuming 25% normal DNA contamination is common in prostate cancers. Blue, normal DNA; Pink, tumor DNA. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

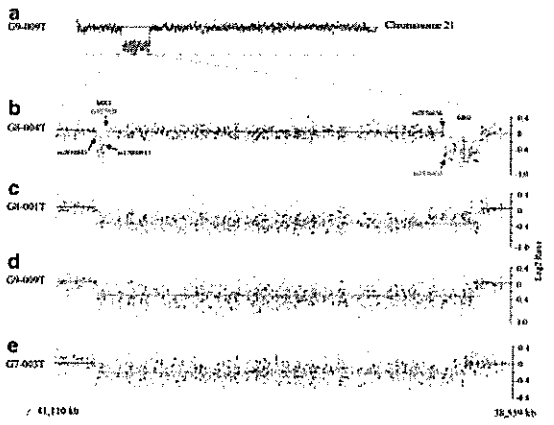

Figure 2. Scatter plot of raw log2 ratios for the SNP probes between *TMPRSS2* and *ERG*. (a) log2 ratio of nonallele based analysis of chromosome 21 with 10-SNP genomic smoothing in G9-009T showing the relative position of the deleted region to be analyzed. (b–e) Raw log2 ratios of SNP probes between *TMPRSS2* and *ERG*. The genomic DNA between the SNPs rs457920 and rs2836656 was apparently retained in G8-004T, while at least one copy of this region was deleted in G8-001 (c), G9-009T (d), and G7-003T (e). Blue dots, raw data of log2 ratios from Nsp array. Pink dots, raw data of log2 ratios from Sty array. The *y*-axis represents the log2 ratio obtained using CNAG2.0. The arrows point to SNPs that map to the regions containing the breakpoints of the deletions. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

in Figure 1, including G6-002T, G6-003T, G7-002T, G7-003T, G7-013T, G7-016T, G7-017T, G7-024T, G7-029T, G7-032T, G8-002T, G8-003T, and G9-004T. The results from independent assays using the Nsp array and the Sty array were very consistent (Figs. 1a–1d), and this consistency was used to confirm the genomic loss of one copy in this region. While most of the hemizygous deletions among these 13 tumors appeared to be similar, as shown in Figures 1a–1r, two of them (Figs. 1f and 1q) were apparently different in terms of the size and break point positions of the deletions.

Among these 16 tumors with T-E deletions, three appeared to harbor genomic losses that affected both of the alleles of *TMPRSS2*, *ERG*, and other genes in between (Fig. 1m, G9-009T; Fig. 1n, G8-001T; and Fig. 1q, G8-004T). Analyzing the raw data of log2 ratio as shown in Figure 2, we found that the homozygous deletion in G8-004T also affected *MX1* in addition to *TMPRSS* and *ERG*. We confirmed the homozygous deletions using qPCR as shown in Figures 1s–1u. All of the other genes between *TMPRSS2* and *ERG*, including *ETS2*, *DSCR2*, *BRWD1*, *HMGN1*, *SH3BGR*, *WRB*, *B3GALT5*, *PCP4*, *DSCAM*, *BACE2*, *PLAC4*, *FAM3B*, and *MX2*, were apparently retained in the genome of G8-004T. The raw data of log2 ratio also revealed different break points involving both *TMPRSS2* and *ERG* in different tumors (Fig. 2), which indicates that different breakpoint positions may contribute to the variety of fusions of *TMPRSS2* and *ERG* observed in different prostate tumors.

DNA Breakpoints Involved in Different Regions Within both *TMPRSS2* and *ERG* Genes

In order to assess the contribution of the genomic regions harboring these breakpoints to the marked differences in the transcripts of the fusion gene, we analyzed the raw probe intensity

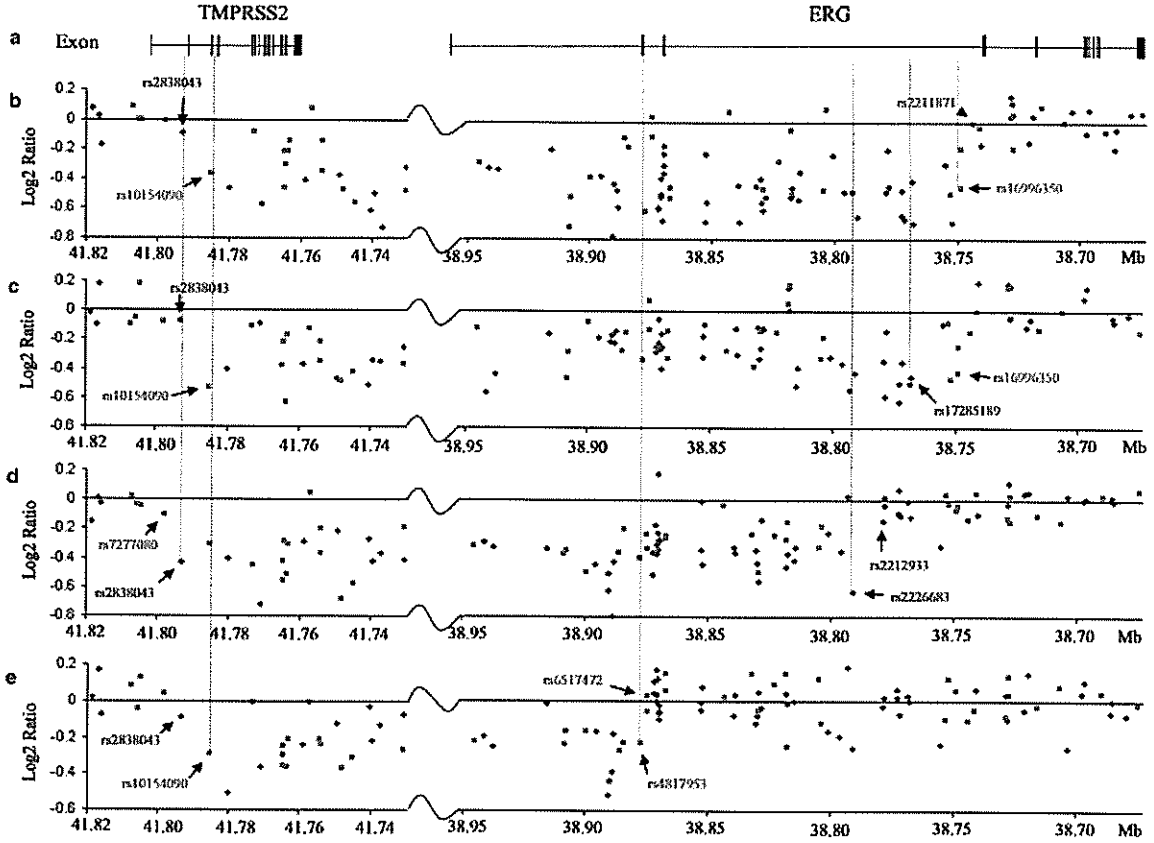

Figure 3. Mapping the breakpoints of genomic deletions in different tumors to various introns of *TMPRSS2* and *ERG*. (a) Relative positions of the exons and introns in *TMPRSS2* and *ERG* (UCSC hg17, NM_005656.2, MN_004449.3). (b–e) The raw data of log2 ratios, showing the regions involved in DNA copy number changes that harbor the breakpoints of genomic deletions, as well as the SNP probes that map to these regions: (b) G8-004T, (c) G8-001T, (d) G9-009T, and (e) G7-003T. Blue diamonds, raw data of log2 ratios from Nsp array. Pink squares, raw data of log2 ratios from Sty array. The *y*-axis represents the log2 ratio obtained using CNAG2.0, and the *x*-axis represents the physical position of the SNPs (UCSC hg17). The arrows point to the SNPs that map to the regions containing the breakpoints of the deletions. The vertical dotted lines depict the introns where SNPs locate. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

data of log2 ratios generated from CNAG2.0 for the tumors with the deletions described above, and present some examples of the results in Figure 3. We also used a 3-SNP genomic smoothing approach to reduce the noise of the raw data in order to identify the breakpoint SNPs as presented in Supplementary Table 2 (http://www1.wfubmc.edu/Genomics/Publications+and+Data). In the tumor G9-009T (Fig. 3d), a change of log2 ratio in *TMPRSS2* apparently occurred between the first and the second exons, while an apparent shift of log2 ratio occurred between exons 3 and 4 in *ERG* (marked by vertical dote lines). In G7-003T (Fig. 3e) the obvious decrease of log2 ratios occurred between exons 2 and 3 in *ERG*, while an apparent shift of log2 ratio occurred between exons 1 and 4 in *TMPRSS2*. In G8-004T and G8-001T (Figs. 3b and 3c, respectively), the changes of log2 ratios in *ERG* clearly occurred between exons 3 and 4, while the shift of log2 ratio was observed between exons 1 and 4 in *TMPRSS2*. By analyzing both the raw and smoothed hybridization intensity of probes at each SNP between *TMPRSS2* and *ERG*, we found that the deletions started and ended at various locations in different tumors (Figs. 1–3).

Breakpoints of Genomic Deletions at Multiple Locations Correlate with Various Forms of Fusion Transcripts

The high resolution of the 500K SNP array enabled us to map the regions where the deletions started and ended to the introns within *TMPRSS2* and *ERG* for most of the tumors. Therefore, it was possible for us to predict the structure of the potential transcripts that would result from the fusion genes. For example, in tumor G9-009T the deletion between SNPs rs2838043 and rs2226683 (Fig. 3d) would create a fusion between the first

TABLE 1. Expression of *TMPRSS/ERG* Fusion Transcripts in Human Prostate Tumors

Subject	T1/*ERG*	T1-2/*ERG*	*ERG* v1	*ERG* v2
Vcap	T1/E4-11		+	+
N1 (650N)			+	
1647 lbbh1			+	
1000 aaa	T1/E4-11		+	+
G7-003T		T1-2/E3-11	+	+
G7-004T	T1/E4-11	T1-2/E4-11	+	
G7-029T	T1/E4-11		+	+
G7-016T	T1/E4-11		+	+
G7-042T	T1/E4-11	T1-2/E4-11	+	+
G8-003T	T1/E4-11		+	+
G8-004T	T1/E4-11	T1-2/E4-11	+	+
G9-004T	T1/E4-11		+	+
G9-005T			+	
G9-007T	T1/E4-11	T1-2/E4-11	+	+
G9-008T	T1/E4-11		+	+
G9-010T		T1-2/E4-11	+	

Vcap, cell line; N1 (650N), nonmalignant tissues; 1000aaa, positive (+) control for T1/E4-11; 1647 lbbh1, positive (+) control for T1-2/E4-11; *ERG* v1, *ERG* variant 1 (NM_182918); *ERG* v2, *ERG* variant 2 (NM_004449.3).

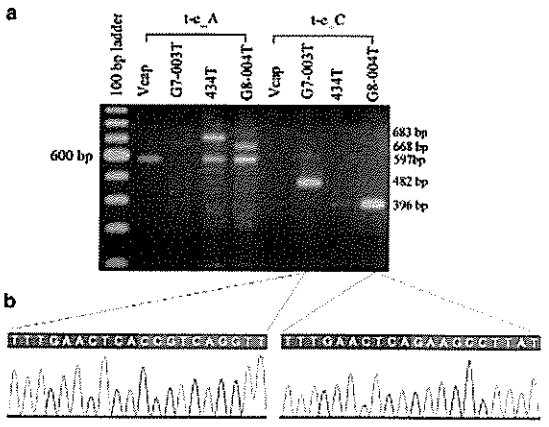

Figure 4. Analysis of *TMPRSS2/ERG* transcripts in different prostate tumors. (a) RT-PCR products resolved in 1.5% agarose gel. t-e_ A primers: *TMPRSS2* (cDNA NM_005656) 12-28: CAGGAGGCGGAGGCGGA, *ERG* (cDNA NM_004449) 762-742: GGCGTTGT A GCTGGGGGTGAG; t-e_C primers: *TMPRSS2* (cDNA NM_005656) 121–141: CAGCAAGATGGCTTTGAACTC, *ERG* (cDNA NM_004449) 599–580: GGATCTGCTGGCACGATAAC. Vcap, celline as positive control. (b) Sequence analysis of the fusion transcripts in G7-003 and G8-004. Blue horizontal bar, *TMPRSS2*. Red horizontal bar, *ERG*. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

intron of *TMPRSS2* and the third intron of *ERG*, which would result in the most frequent fusion transcript of T1/E4–11 (Fig. 5) as described previously (Tomlins et al., 2005; Wang et al., 2006; Clark et al., 2007). This was confirmed by analysis of the fusion transcripts in the tumors G9-008T, G9-007T, G9-004T, G8-004T, G8-003T, G7-016T, and G7-029T (Table 1). In G7-003T (Fig. 3e), the deletion between the SNPs rs4817953 and rs10154090 would be expected to create a fusion between the second intron of *ERG* and 1st, 2nd, or the 3rd intron of *TMPRSS2*. Because the region that harbors the deletion breakpoint lies between the two SNPs (rs2838043 and rs10154090) and thus covers three introns, with the 2nd and the 3rd exons of *TMPRSS2* in between, it was impossible for us to map the breakpoint to a narrower region based upon the currently available resolution of the 500K SNP array. Nevertheless, the fusion of genomic DNA via this deletion could create a fusion gene producing transcripts T1/E3–11, T1–2/E3–11 or T1–3/E3–11, depending on where the breakpoint occurs in *TMPRSS2*. To confirm the predicted transcript which would result based upon the genomic deletion, we analyzed the fusion transcript in this tumor using two sets of RT-PCR primers. The first set, t-e_A, is located in the 1st exon of *TMPRSS2* and the 6th exon of *ERG*. The second set, t-e_C, is located in the 2nd exon of *TMPRSS2* and the 6th exon of *ERG*. The size and the sequence of the PCR product confirmed that T1–2/E3–11, a new form of fusion transcript, was produced from the fusion gene in this tumor (Fig. 4). In G8-004T (Fig. 3b), the deletion between SNPs rs16996350 and rs10154090 would be expected to create a fusion between the 3rd intron of *ERG* and the 1st, 2nd, or 3rd intron of *TMPRSS2*. While we could not map the breakpoint to a narrower region for the same reasons described above, the fusion of genomic DNA via this deletion could create a gene producing fusion transcripts consisting of T1/E4–11, T1–2/E4–11, or T1–3/E4–11. Size and sequence analyses of the transcripts revealed that both T1/E4–11 and T1–2/E4–11 were produced in this tumor (Fig. 4). Therefore, the deletions with breakpoints mapped to different locations in various introns within *TMPRSS2* and *ERG* as presented in Figure 3 contributed to the production of different forms of the fusion transcripts observed in various prostate tumors.

Expression of Fusion Transcripts Correlated with Transcription Variant 2 of *ERG*

There are two transcript variants of *ERG* according to UCSC RefSeq Gene database (NM_182918, NM_004449). We analyzed the expression of these two variants and various *TMPRSS2/ERG* fusion transcripts in a subset of the prostate tumors for which RNA was available using RT-PCR (Table 1). The RT-PCR products confirmed the fusion of both T1 and T1–2 to different exons of *ERG* by sequencing analysis (Fig. 4). It is interesting that the expression of *ERG* variant 2 was only observed

in tumors expressing the fusion transcripts, whereas the expression of variant 1 was detected in all samples. It is also worth noting that some of the fusion transcripts were found in tumors (e.g., G9-007T, G9-008T, G7-042T) where T-E deletion was not detected by the 500K SNP array (Table 1).

DISCUSSION

Two major findings have been reported since the discovery of the recurrent fusion between *TMPRSS2* and *ERG* transcription factor genes in prostate cancer by Tomlins et al. (2005). First, we and others have demonstrated that genomic deletions contribute to the fusion between *TMPSS2* and *ERG* (Liu et al., 2006; Perner et al., 2006; Yoshimoto et al., 2006). Second, a number of distinct fusion transcripts have been reported with some of them associated with the outcome of prostate cancer (Clark et al., 2006; Soller et al., 2006; Wang et al., 2006; Demichelis et al., 2007). Although alternative splicing of an mRNA encoded by a single *TMPRSS2/ERG* fusion gene has been proposed for this phenomenon (Wang et al., 2006; Clark et al., 2007), the source of the diversity observed among these fusion transcripts remains elusive. In our new work, we now demonstrate that genomic alterations at different locations within *TMPRSS2* and *ERG* contribute to the production of various fusion transcripts found in various prostate cancer cells, based on our data showing an association between a specific genomic deletion and a particular form of fusion transcript in several tumors. This conclusion is supported by the observation of multiple breakpoints in introns 1 and 2 of *TMPRSS2* and corresponding transcripts in xenograft DNAs (Hermans et al., 2006).

Using allele-specific analysis of the allele intensity and genotype data generated from the GeneChip 500K SNP array, we first showed that both hemizygous and homozygous deletions occurred between and within *TMPRSS2* and *ERG* in prostate tumors. In tumor G8-004T, with homozygous deletions of 8 of the 11 exons in *ERG* and at least 12 of the 14 exons in *TMPRSS2*, we found that all of the genes between *TMPRSS2* and *ERG* remained intact except *MX1*, which lost 11 of its 17 exons (Fig. 2b). This result suggests that deletions of these genes may not be as biologically important as the fusion of *TMPRSS2* and *ERG* in terms of prostate cancer progression. Although it is not clear whether the nondeleted portion still remains on chromosome 21, the presence of the fusion transcripts, T1/E4–11 and T1–2/E4–11, in this tumor indicates a portion of the DNA between these two genes may have translocated to another location in the genome. Therefore, in addition to the deletion, translocation is another possible mechanism mediating the fusion of *TMPRSS2* and *ERG*, which is consistent with the observations from previous studies (Tomlins et al., 2005; Hermans et al., 2006).

In the other 15 tumors harboring T-E deletions, at least one copy of these 14 known genes described above was lost (Fig. 1). Although the loss and decreased expression of the genes between *TMPRSS2* and *ERG* have been proposed to be associated with cancer progression and tumor growth respectively (Birger et al., 2005; Perner et al., 2006), the synergetic effects of the loss of these genes and the fusion of *TMPRSS2* and *ERG* on cancer progression need to be further investigated in a larger population of prostate tumors before definitive conclusions can be drawn.

To further characterize the regions involved in the deletions, we aligned the regions where the breakpoints were mapped using CLC Free Workbench bioinformatics software (http://clcbio.com). It is interesting that there is at least one consensus sequence identified in the involved regions of the introns of *TMPRSS2* and *ERG* with ~80% of the two sequences identical. These results are consistent with the findings of Yoshimoto et al. (2006) who reported that at least one area in the intron following the transcribed *TMPRSS2* exon had up to 90% homology to multiple areas of the intron preceding the relevant *ERG* exon for each of the fusion transcripts in their study. We further found that these consensus sequences exist not only in *TMPRSS2* and *ERG* introns that are involved in partial deletions between these two genes, but also in other rearrangements, e.g. in the *ERG* intron and a nongenic region that are involved in a deletion of 8 *ERG* exons and in the *MX1* and *TMPRSS2* introns that are involved in a deletion of 11 *MX1* exons and at least 12 *TMPRSS2* exons without deleting the other genes located in between (see Fig. 2b). When aligning the two consensus sequences identified in the regions containing the breakpoints of the deletion, we found that they are very similar to the human Alu-Sq and Alu-Sp subfamily consensus sequences, with more than 80% homology.

Genomic deletions mediated by recombination between Alu elements in humans have been reported. Some of the genomic alterations associated with Alu repeats have been observed to be correlated with various cancers including heredi-

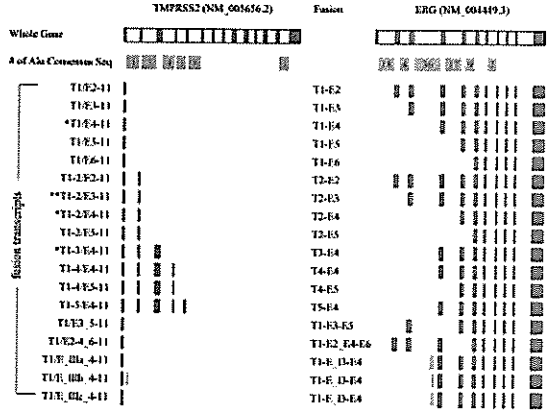

Figure 5. Schematic presentation of the relationship between the locations of the human Alu family consensus sequence in *TMPRSS* and *ERG* and the forms of the fusion transcripts summarized by Clark et al. (2007). *Marks the *TMPRSS2/ERG* fusion transcripts identified/predicted in the tumors analyzed in this study. The sizes of the introns and exons of these two genes are not drawn to scale. Dark blue, *TMPRSS2* exons. Light blue, *TMPRSS2* intron containing Alu family consensus sequence. Red, *ERG* exons. Pink, *ERG* introns containing Alu family consensus sequence. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

tary nonpolyposis colorectal and endometrial cancers (Mauillon et al., 1996), acute myeloid leukemia (Strout et al., 1998), and breast cancer (Gad et al., 2002). To explore the association between Alu distribution and the diversity of *TMPRSS2/ERG* fusion transcripts in prostate cancers, we searched the human Alu family consensus sequence in *TMPRSS2* and *ERG* using the reference genes (UCSC hg17, NM_005656.2, MN_004449.3) in the NCBI Bl2seq utility. As presented in Figure 5, the distribution of Alu matches well with all 17 structures of *TMPRSS2/ERG* fusion transcripts reported so far (summarized by Clark et al., 2007) and the new fusion transcript identified in this study, where the retaining of exons 1–5 of *TMPRSS2* and deletion of exons 1–5 of *ERG* were observed in various tumors. The presence/absence of Alu family consensus sequence in the introns of *TMPRSS2* and *ERG* apparently correlates with the presence/absence of fusion transcripts of theses two genes. Therefore, the Alu element may indirectly contribute to the diversity of the fusion transcripts found in various prostate cancers by facilitating recombination for deletion, translocation, or/and fusion of *TMPRSS2* and *ERG*. If this is the case, additional new forms of fusion transcripts will be identified when more prostate tumors are analyzed in future studies. On the other hand, the association between presence/absence of fusion transcripts and presence/absence of Alu elements as shown in Figure 5 could be due to chance, because of the abundance of Alu sequences in the human genome (more than one million copies). Therefore, whether these consensus sequences mediate recombination for deletion, translocation or/and fusion of *TMPRSS2* and *ERG* in prostate cancer needs to be further investigated.

We also found a significant association (P = 0.0179, Fisher's exact test) between the expression of the *TMPRSS2/ERG* fusion transcript and the expression of the *ERG* variant 2 transcript (NM_004449.3) that lacks an in-frame exon compared to variant 1 (NM_182918) in prostate tumors (Table 1), although the molecular mechanism remains unknown. However, this observation should be confirmed using a large population from a different cohort in the future studies.

It is worth noting that detectable deletions/translocations at chromosome band 21q22 cannot account for all of the fusion transcripts found in various tumors. For example, some of the fusion transcripts were also detected in the tumors (G7-042T, G9-007T, G9-008T) where apparent T-E deletion was not detected by the 500K SNP array (Table 1). While the process that produced these transcripts remains unknown, they might result from balanced translations which can be further assessed in these types of tumor cells using different techniques, such as FISH and chromosomal painting.

The correlation between the expression of various fusion transcripts and aggressiveness/clinical outcome of prostate cancer has been contradictory across various studies. Although the expression of some *TMPRSS2/ERG* transcripts was reported to be associated with clinical outcome of prostate cancer (Parner et al., 2006; Wang et al., 2006), we found no significant correlation between T-E deletion and Gleason score (from Gleasons 6 to 9) in our study population. This result is consistent with the observations reported by Yoshimoto et al. (2006) and Lapointe et al. (2007). However, Winnes et al. (2007) observed a tendency of fusion-positive tumors being associated with lower Gleason grade and better survival than fusion-negative tumors. On the other hand, Demichelis et al. (2007), Mehra et al. (2007), Nam et al. (2007), and Rajput et al. (in press) reported that *TMPRSS2/ERG* fusion was associated with lethal, high pathologic-stage, higher rate of recurrence and poorly differentiated tumors, respectively, providing thus far the strongest correlation between the fusion of these two genes and the outcome of prostate cancers.

In summary, we have demonstrated that both hemizygous and homozygous deletions occurred

between and within *TMPRSS2* and *ERG* at different locations on chromosome band 21q22 in prostate tumors. The presence of specific genomic deletions between *TMPRSS2* and *ERG* and a corresponding form(s) of fusion transcript(s) in various tumors demonstrates that multiple genomic alterations/translocations contribute to the diversity of the fusion transcripts found in various prostate cancers.

ACKNOWLEDGMENTS

The authors thank all the study subjects who participated in this study.

REFERENCES

Birger Y, Catez F, Furusawa T, Lim JH, Prymakowska-Bosak M, West KL, Postnikov YV, Haines DC, Bustin M. 2005. Increased tumorigenicity and sensitivity to ionizing radiation upon loss of chromosomal protein HMGN1. Cancer Res 65:6711–67188.

Bova GS, MacGrogan D, Levy A, Pin SS, Bookstein R, Isaacs WB. 1996. Physical mapping of chromosome 8p22 markers and their homozygous deletion in a metastatic prostate cancer. Genomics 35:46–54.

Clark J, Merson S, Jhavar S, Flohr P, Edwards S, Foster CS, Eeles R, Martin FL, Phillips DH, Crundwell M, Christmas T, Thompson A, Fisher C, Kovacs G, Cooper CS. 2007. Diversity of *TMPRSS2-ERG* fusion transcripts in the human prostate. Oncogene 19; 26:2667–2673.

Cerveira N, Ribeiro FR, Peixoto A, Costa V, Henrique R, Jeronimo C, Teixeira MR. 2006. *TMPRSS2-ERG* gene fusion causing *ERG* overexpression precedes chromosome copy number changes in prostate carcinomas and paired HGPIN lesions. Neoplasia 8:826–832.

Demichelis F, Fall K, Perner S, Andren O, Schmidt F, Setlur SR, Hoshida Y, Mosquera JM, Pawitan Y, Lee C, Adami HO, Mucci LA, Kantoff PW, Andersson SO, Chinnaiyan AM, Johansson JE, Rubin MA. 2007. *TMPRSS2:ERG* gene fusion associated with lethal prostate cancer in a watchful waiting cohort. Oncogene 26:4596–4599.

Gad S, Caux-Moncoutier V, Pages-Berhouet S, Gauthier-Villars M, Coupier I, Pujol P, Frenay M, Gilbert B, Maugard C, Bignon YJ, Chevrier A, Rossi A, Fricker JP, Nguyen TD, Demange L, Aurias A, Bensimon A, Stoppa-Lyonnet D. 2002. Significant contribution of large BRCA1 gene rearrangements in 120 French breast and ovarian cancer families. Oncogene 21:6841–6847.

Hermans KG, van Marion R, van Dekken H, Jenster G, van Weerden WM, Trapman J. 2006. *TMPRSS2:ERG* fusion by translocation or interstitial deletion is highly relevant in androgen-dependent prostate cancer, but is bypassed in late-stage androgen receptor-negative prostate cancer. Cancer Res 66:10658–10663.

Lapointe J, Kim YH, Miller MA, Li C, Kaygusuz G, van de Rijn M, Huntsman DG, Brooks JD, Pollack JR. 2007. A variant TMPRSS2 isoform and ERG fusion product in prostate cancer with implications for molecular diagnosis. Mod Pathol 20:467–473.

Lin M, Wei LJ, Sellers WR, Lieberfarb M, Wong WH, Li C. 2004. dChipSNP: Significance curve and clustering of SNP-array-based loss-of-heterozygosity data. Bioinformatics 20:1233–1240.

Liu W, Chang B, Sauvageot J, Dimitrov L, Gielzak M, Li T, Yan G, Sun J, Sun J, Adams TS, Turner AR, Kim JW, Meyers DA, Zheng SL, Isaacs WB, Xu J. 2006. Comprehensive assessment of DNA copy number alterations in human prostate cancers using Affymetrix 100K SNP mapping array. Genes Chromosomes Cancer 45:1018–1032.

Mauillon JL, Michel P, Limacher J, Latouche J, Dechelotte P, Charbonnier F, Martin C, Moreau V, Metayer J, Paillot B, Frebourg T. 1996. Identification of novel germline hMLH1 mutations including a 22 kb Alu-mediated deletion in patients with familial colorectal cancer. Cancer Res 56:5728–5733.

Mehra R, Tomlins SA, Shen R, Nadeem O, Wang L, Wei JT, Pienta KJ, Ghosh D, Rubin MA, Chinnaiyan AM, Shah RB. 2007. Comprehensive assessment of TMPRSS2 and ETS family gene aberrations in clinically localized prostate cancer. Mod Pathol 20:538–544.

Nami RK, Sugar L, Wang Z, Yang W, Kitching R, Klotz LH, Venkateswaran V, Narod SA, Seth A. 2007. Expression of TMPRSS2 ERG gene fusion in prostate cancer cells is an important prognostic factor for cancer progression. Cancer Biol Ther 6:40–45.

Nannya Y, Sanada M, Nakazaki K, Hosoya N, Wang L, Hangaishi A, Kurokawa M, Chiba S, Bailey DK, Kennedy GC, Ogawa S. 2005. A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays. Cancer Res 15:6071–6079.

Perner S, Demichelis F, Beroukhim R, Schmidt FH, Mosquera JM, Setlur S, Tchinda J, Tomlins SA, Hofer MD, Pienta KG, Kuefer R, Vessella R, Sun XW, Meyerson M, Lee C, Sellers WR, Chinnaiyan AM, Rubin MA. 2006. *TMPRSS2:ERG* fusion-associated deletions provide insight into the heterogeneity of prostate cancer. Cancer Res 66:8337–8341.

Petrovics G, Liu A, Shaheduzzaman S, Furasato B, Sun C, Chen Y, Nau M, Ravindranath L, Chen Y, Dobi A, Srikantan V, Sesterhenn IA, McLeod DG, Vahey M, Moul JW, Srivastava S. 2005. Frequent overexpression of ETS-related gene-1 (ERG1) in prostate cancer transcriptome. Oncogene 24:3847–3852.

Rajput AB, Miller MA, De Luca A, Boyd N, Leung S, Hurtado-Coll A, Fazli L, Jones EC, Palmer JB, Gleave ME, Cox ME, Huntsman DG. Frequency of the *TMPRSS2:ERG* gene fusion is increased in moderate to poorly differentiated prostate cancers. J Clin Pathol (in press).

Rubinstein YR, Furusawa T, Lim JH, Postnikov YV, West KL, Birger Y, Lee S, Nguyen P, Trepel JB, Bustin M. 2005. Chromosomal protein HMGN1 modulates the expression of N-cadherin. FEBS J 272:5853–5863.

Soller MJ, Isaksson M, Elfving P, Soller W, Lundgren R, Panagopoulos I. 2006. Confirmation of the high frequency of the TMPRSS2/ERG fusion gene in prostate cancer. Genes Chromosomes Cancer 45:717–719.

Strout MP, Marcucci G, Bloomfield CD, Caligiuri MA. 1998. The partial tandem duplication of ALL1 (MLL) is consistently generated by Alu-mediated homologous recombination in acute myeloid leukemia. Proc Natl Acad Sci USA 95:2390–2395.

Tomita K, van Bokhoven A, van Leenders GJLH, Ruijter ETG, Jansen CFJ, Bussemakers MJG, Schalken JA. 2000. Cadherin switching in human prostate cancer progression. Cancer Res 60:3650–3654.

Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, Varambally S, Cao X, Tchinda J, Kuefer R, Lee C, Montie JE, Shah RB, Pienta KJ, Rubin MA, Chinnaiyan AM. 2005. Recurrent fusion of *TMPRSS2* and ETS transcription factor genes in prostate cancer. Science 310:644–648.

Wang J, Cai Y, Ren C, Ittmann M. 2006. Expression of variant *TMPRSS2/ERG* fusion messenger RNAs is associated with aggressive prostate cancer. Cancer Res 66:8347–8351

Winnes M, Lissbrant E, Damber JE, Stenman G. 2007. Molecular genetic analyses of the TMPRSS2–ERG and TMPRSS2–ETV1 gene fusions in 50 cases of prostate cancer. Oncol Rep 17:1033–1036.

Yoshimoto M, Joshua AM, Chilton-Macneill S, Bayani J, Selvarajah S, Evans AJ, Zielenska M, Squire JA. 2006. Three-color FISH analysis of *TMPRSS2/ERG* fusions in prostate cancer indicates that genomic microdeletion of chromosome 21 is associated with rearrangement. Neoplasia 8:465–469.

Integration of Somatic Deletion Analysis of Prostate Cancers and Germline Linkage Analysis of Prostate Cancer Families Reveals Two Small Consensus Regions for Prostate Cancer Genes at 8p

Bao-li Chang,[1] Wennuan Liu,[1] Jishan Sun,[1] Latchezar Dimitrov,[1] Tao Li,[1] Aubrey R. Turner,[1] Siqun L. Zheng,[1] William B. Isaacs,[2] and Jianfeng Xu[1]

[1]Center for Human Genomics, Wake Forest University School of Medicine, Winston-Salem, North Carolina and [2]Johns Hopkins Medical Institutions, Baltimore, Maryland

Abstract

The evidence for tumor suppressor genes at 8p is well supported by many somatic deletion studies and genetic linkage studies. However, it remains a challenge to pinpoint the tumor suppressor genes at 8p primarily because the implicated regions are broad. In this study, we attempted to narrow down the implicated regions by incorporating evidence from both somatic and germline studies. Using high-resolution Affymetrix arrays, we identified two small common deleted regions among 55 prostate tumors at 8p23.1 (9.8–11.5 Mb) and 8p21.3 (20.6–23.7 Mb). Interestingly, our fine mapping linkage analysis at 8p among 206 hereditary prostate cancer families also provided evidence for linkage at these two regions at 8p23.1 (5.8–11.2 Mb) and at 8p21.3 (19.6–23.9 Mb). More importantly, by combining the results from the somatic deletion analysis and genetic linkage analysis, we were able to further narrow the regions to ~1.4 Mb at 8p23.1 and ~3.1 Mb at 8p21.3. These smaller consensus regions may facilitate a more effective search for prostate cancer genes at 8p. [Cancer Res 2007;67(9):4098–103]

Introduction

Deletion of sequences from chromosome 8p is the most common deletion event in the genome of prostate tumors (1). In a recent study that estimated the frequency of DNA copy number alterations in the prostate cancer genome based upon all published comparative genomic hybridization studies of prostate cancers, we found that one third of 891 prostate cancers had a deletion at 8p21.3, considerably higher than the second most commonly deleted region at 6q15 (22.4%; ref. 2). Despite the overwhelming evidence for 8p deletions, few specific genes have been consistently implicated as prostate tumor suppressor genes in this region. One of the major obstacles in the identification of tumor suppressor genes at 8p is the size of the deleted regions, which is affected by the resolution of methods used to detect deletions. For example, in our study cited above (2), the deleted region at 8p21.3 spans a 27.1-Mb interval extending into 8p23.3 and 8p21.1 and contains many genes. Higher-resolution detection methods that can detect small deletions and complex deletion patterns are needed to identify 8p tumor suppressor genes (3).

Furthermore, results from genetic linkage studies have provided evidence for prostate cancer linkages at 8p (4, 5). In the largest genome-wide linkage analysis done to date, Xu et al. (5) found suggestive evidence for linkage at 8p21, one of the five most significant in the genome, among 1,233 prostate cancer families of the International Consortium for Prostate Cancer Genetics (ICPCG). Similar to the results of somatic deletion studies, few genes in this 8p region have been consistently implicated as major prostate cancer susceptibility genes accounting for the 8p linkage. One of the major difficulties is the low resolution of genetic linkage studies, which are typically in the range of 10 to 20 cM, due to limited meiosis events in families. For example, the 1-LOD drop interval of 8p21 linkage identified in the ICPCG study was 13 cM (39–52 cM) or 10 Mb (22–32 Mb).

Cancers are thought to arise as a result of alterations in expression of tumor suppressor genes and oncogenes in prostate epithelial cells. Altered gene expression may result from inherited genetic changes and acquired somatic genetic changes, including deletions, as hypothesized by the "two-hit" model (6). Therefore, assuming that at least some fraction of prostate cancers arise from a combination of inherited and acquired genomic events affecting the same gene or combination of genes, studies that simultaneously examine inherited genetic changes and somatic genetic alterations of chromosomal regions or genes may improve the likelihood of identifying genes involved in cancer development.

In this study, we have taken three steps to identify genomic regions that contain prostate cancer genes. First, we used high-resolution Affymetrix single nucleotide polymorphism (SNP) arrays to examine detailed deletion patterns at 8p among 55 prostate cancers. Second, we did a fine mapping linkage analysis in 206 hereditary prostate cancer (HPC) families. Finally and more importantly, we integrated results from somatic deletion analysis and germline linkage analysis to identify a consensus region.

Materials and Methods

Detection of 8p deletions in somatic DNA from prostate cancers. All subjects were prostate cancer patients undergoing radical prostatectomy for treatment of clinically localized disease at Johns Hopkins Hospital. For the somatic DNA deletion analysis of prostate cancers, we selected 55 subjects from whom genomic DNA of sufficient quantity (>5 μg) and purity (>70% cancer cells for cancer specimens, no detectable cancer cells for normal samples) could be obtained by macrodissection of matched nonmalignant (hereafter referred to as "normal") and cancer containing areas of prostate tissue as determined by histologic evaluation of H&E-stained frozen sections of snap-frozen radical prostatectomy specimens.

Note: B-L. Chang and W. Liu contributed equally to this study.

Requests for reprints: William B. Isaacs, Johns Hopkins Hospital, Marburg 115, 600 North Wolfe Street, Baltimore, MD 21287. Phone: 410-955-2518; Fax: 410-955-0833; E-mail: wisaacs@jhmi.edu or Jianfeng Xu, Center for Human Genomics, Medical Center Boulevard, Winston-Salem, NC 27157. Phone: 336-713-7500; Fax: 336-713-7566; E-mail: jxu@wfubmc.edu.

©2007 American Association for Cancer Research.

doi:10.1158/0008-5472.CAN-06-4570

Genomic DNA was isolated from trimmed frozen tissues as previously described (7).

We used Affymetrix SNP array panels to detect DNA copy number alterations, and for this study, we focused entirely on 8p deletions. We used the 100K SNP array for the first 22 subjects (3) and then used the 500K SNP array for the final 41 subjects (eight samples were analyzed using both 100K and 500K arrays). For all subjects, we analyzed both tumor DNA and normal DNA from the same subject using Affymetrix SNP arrays (either 100K or 500K array). The normal DNAs were extracted from histologically normal prostate tissue from the same prostate or from the seminal vesicle of the same patient. DNA copy number was calculated based on allele intensity using two different software packages: Copy Number Analyzer for Affymetrix GeneChip (CNAG2.0; ref. 8) and dChip analyzer (9). Allele-specific analysis was also done to estimate DNA copy number for each chromosome using CNAG2.0. The physical positions of detected deletions were based on the Human hg17 Assembly (NCBI Build 35). The criteria used in dChip analysis for this study are similar to those used for the 100K SNP array analyses using CNAT, which has been described in our previous publication (3). Briefly, to reduce random noise in allele intensity at individual SNPs, we first estimated DNA copy number based on flanking SNPs in the region, using the 10-SNP smoothing setting of dChip software to obtain a genome smooth average copy number (GSACN) for each SNP. We then defined deletions using the following working criteria: a minimum of four consecutive SNPs with at least three of them having the following characteristics: the GSACN ratios of tumor/matched normal <0.75, the GSACN of the tumor DNA <1.9, and the minimum physical length of the putative deletion ≥2 kb. To define deletions using CNAG2.0 software (for both intensity-based and allele-specific analyses), we also used a 10-SNP smoothing setting to minimize random variations at individual SNPs. Each deletion is defined by whether the log 2 ratio of probe intensity is below

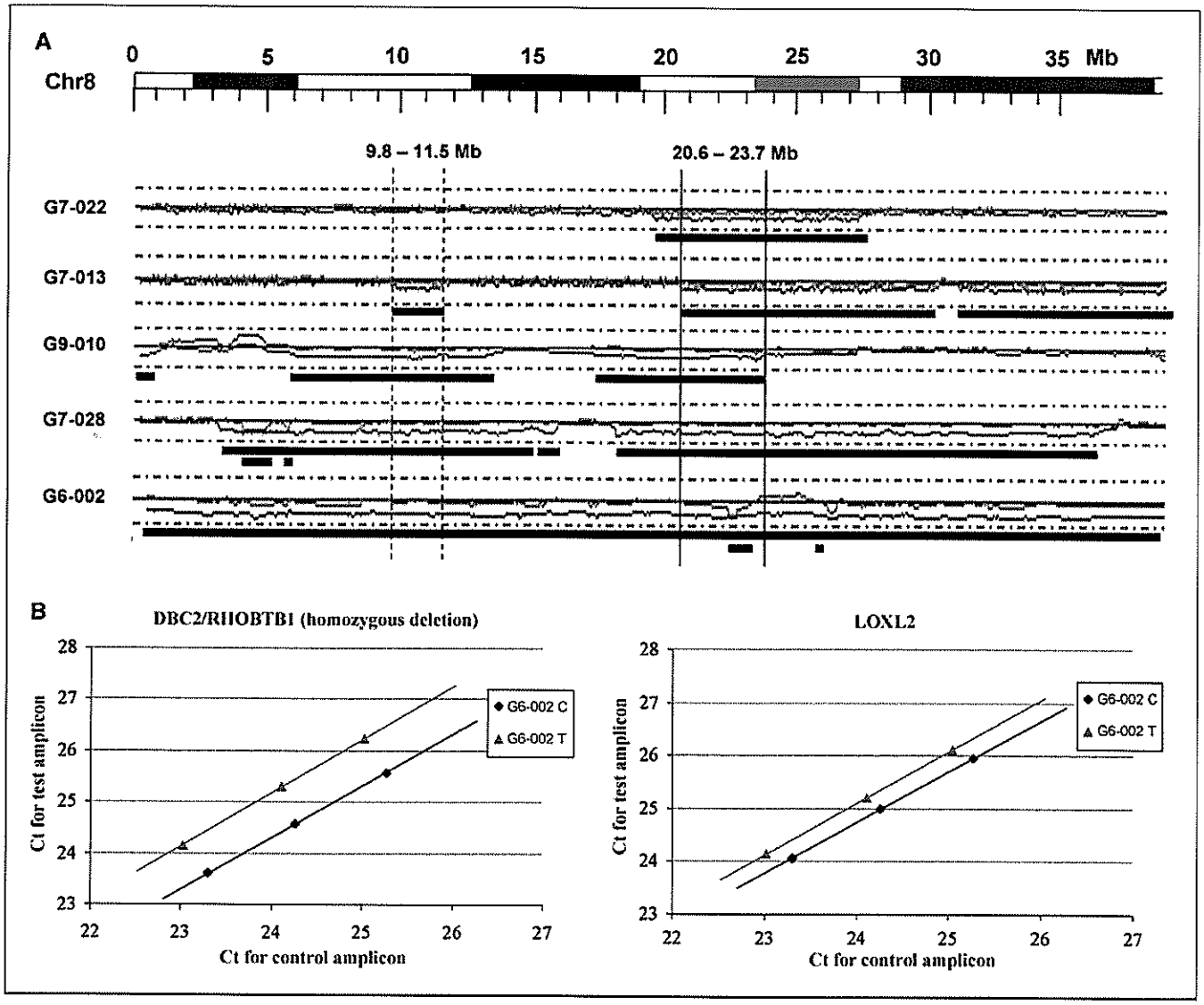

Figure 1. Small and complex partial deletions of 8p. *A*, a visual summary for five of the smallest and most complex tumors that were detected at 8p. *Top*, distances in Mb beginning at the pter (*Chr8*). Allele-specific CNAG output for each of these five tumors, with red or green for each respective chromosome. Solid vertical lines highlight a region at 8p21.3-8p21.2, spanning from 20.6 to 23.7 Mb, that was deleted in all 30 tumors. Dotted vertical lines indicate a deleted region spanning from 9.8 to 11.5 Mb at 8p23.1 that is shared by 29 of the 30 tumors. *B*, results of deletion confirmation of the homozygous deleted regions by quantittaive real-time PCR analyses. *Left*, results from *DBC2/RHOBTB1* primers located within the putative homozygous deletion. *Right*, results from *LOXL2* primers located outside the homozygous deleted interval. C_t values of the control (*X*-axis) and test (*Y*-axis) amplicons for the three dilutions of each DNA sample were plotted against each other. Offsets between best-fit lines for the samples along the test amplicon axis at 25 C_t of the control amplicon axis are used to infer DNA copy number. Tumor DNA is defined as having a hemizygous deletion when $C_t < 0.68$ and as having homozygous deletion when $C_t > 0.68$, assuming 25% normal DNA contamination.

Table 1. Estimated regions of large deletions at 8p

Tumor	Gleason scores	Position (start–end), Mb	
		Hemizygous deletions	Homozygous deletions
G7-022	6	19.5-27.4	
G7-013	7	9.8-11.5, 20.6-30.2, 31.1-40.8	
G7-028	7	3.2-14.7, 15.2-15.9, 18.1-36.7	4.0-5.0, 5.6-5.8
G9-010	9	0.2-0.9, 5.2-13.6, 16.9-23.7, 24.2-24.4, 39.2 to centromere	
G7-017	7	0.2-26.8	
G9-005	9	6.3-39.0	
G9-003	9	4.1-36.1	
G7-021	7	4.9 to centromere	
G7-026	7	0.2-31.5	
G6-002	6	0.2-42.4	22.4-23.0, 25.9-26.0
G7-029	7	0.2-39.0	
G6-015	6	0.2-42.0	
G9-008	9	0.2-39.9	
G8-005	8	0.2 to centromere	
G8-002	8	0.2-42.2	
G9-004	9	0.2-43.4	
G7-033	7	0.2-43.6	
G7-042	7	0.2-43.2	
G7-051	7	0.2 to centromere*	
G7-016	7	0.2-42.3	
G7-023	7	0.2-43.8	
G7-019	7	0.2-43.8	
G7-004	7	0.2-43.0	
G9-001	9	0.2-43.6	
G9-009	9	0.2-43.8	
G9-011	9	0.2-43.5	
G7-015	7	0.2 to centromere*	
G7-024	7	0.2 to centromere*	
G6-016	6	0.2 to centromere*	

Note: The 10-SNP smoothing log 2 ratio output from CANG 2.0 for the study subjects can be found at http://www1.wfubmc.edu/Genomics/Publications+and+Data/.

*Deletion extends beyond the centromere.

(no overlap) with the baseline log 2 ratio defined by the matched normal DNA. The baseline log 2 ratio has a theoretical value of zero, with small variations due to random noise (as shown in Fig. 1). The results from all three analyses are in agreement with each other, except that the boundaries of the deletions defined by different analyses varied from one to five SNPs.

A subset of putative deletions were subject to confirmation by quantitative real-time PCR using the ABI Prism 7000 Sequence Detection System, as described in detail elsewhere (3).

Linkage analysis in prostate cancer families and construction of a recombinant map. All 206 HPC families were collected and studied at the Brady Urology Institute at Johns Hopkins Hospital (Baltimore, MD) as described previously (4). Prostate cancer diagnosis was verified by medical records for each affected male studied. Age at diagnosis of prostate cancer was confirmed either through medical records or from two other independent sources. The mean age at diagnosis was 64.3 years for the cases in these families. Eighty-four percent of the families are non-Jewish Caucasians, 6.9% are Ashkenazi Jewish, and 8.8% are African Americans.

Thirty fine mapping microsatellite markers spanning about 35 Mb at 8p were genotyped in these 206 HPC families. Following multiplex PCR using fluorescently labeled primers, the resulting PCR fragments were separated using capillary electrophoresis using an ABI 3700 sequencer. Marker allele frequencies were estimated from the 214 independent individuals in the data set. The marker order and distances were primarily based on information available from the MAP-O-MAT web site (10). Four markers were not available in the MAP-O-MAT web site; their order and distances were interpolated from the University of California Santa Cruz (UCSC) Genome Browser.[3] Multipoint linkage analyses were done using both parametric and nonparametric methods implemented by the computer program GENEHUNTER-PLUS (11, 12). For the parametric analysis, the same autosomal-dominant model that was used by Smith et al. (13) was assumed. For the nonparametric analysis, the estimated marker identical by descent (IBD) sharing of alleles for the various affected relative pairs was compared with its expected values under the null hypothesis of no linkage (NPL). A statistical "Z-all" in the program was used (14). Allele sharing LOD scores were then calculated based on the statistical "Z-all" and assigning equal weight to all families using the computer program ASM (12).

Results

Detection of somatic DNA deletions. Detectable deletions at 8p were observed in 29 of the 55 prostate cancers (52.73%) examined in this study (Table 1). Although many of these deletions involved almost the entire short arm of chromosome 8, we detected partial 8p deletions in 10 of these tumors. Among these partial deletions of 8p, five were smaller or more complex (Fig. 1A). This includes a tumor with a small deletion at 8p21 (G7-022), three tumors containing multiple interstitial deletions (G7-013, G9-010, and G7-028), and two tumors with several small homozygous deletions (G7-028 and G6-002). No copy number polymorphisms were detected in these samples, and these homozygous deletions are due to somatic DNA loss.

To independently confirm the ability of our method to detect either heterozygous or homozygous deletion events at 8p21.3, we did quantitative real-time PCR analyses for tumor G6-002 using two primer pairs: one located within the putative homozygous deletion (*DBC2/RHOBTB1*) and the other located outside the homozygous deleted interval (*LOXL2*). The results of this analysis were most consistent with the deletion being homozygous at the *DBC2/RHOBTB1* locus (ΔC_t = 0.90) and being flanked by hemizygous deletions at the *LOXL2* locus (ΔC_t = 0.38; Fig. 1B).

The pattern of deletions observed among the partial 8p deletions suggests the presence of two smaller deletion regions. In particular, a region at 8p21.3-8p21.2, spanning from 20.6 to 23.7 Mb, was deleted in all 29 tumors (Fig. 1A, *solid vertical lines*). We detected another deleted region spanning from 9.8 to 11.5Mb at 8p23.1 that is shared by 28 of the 29 tumors (Fig. 1, *dotted vertical lines*). The primary reason these regions seem to be separated is due to the three tumors with interstitial deletions.

Prostate cancer linkage region. Linkage analysis of 206 prostate cancer families provided evidence for a susceptibility gene at 8p from both parametric (using a dominant model) and nonparametric linkage analyses (Fig. 2). Interestingly, two separate linkage peaks were observed. One peak was found at the marker D8S258 of 8p21.3 (20,411,446), with a LOD score of 2.51 (P = 0.0007) and an NPL score of 3.14. The 1-LOD drop interval spanned ~4 Mb, between 19.6 and 23.9 Mb. The other peak was found at the marker D8S503 of 8p23.1 (9,270,543), with a LOD score of 1.50 (P = 0.009) and an NPL score of 2.72. The 1-LOD drop interval spanned ~5.4 Mb, between 5.8 and 11.2 Mb. There were 49 families

[3] http://genome.ucsc.edu/

with LOD scores > 0.588 ($P_{nominal}$ = 0.05) within the 8pter-8p12 region, with all but two of these families being linked to at least one of the two regions described above. Among these families, 18 families had positive LOD scores across these two regions, 16 families had positive LOD scores only at 8p21.3, and 13 families had positive LOD scores only at 8p23.1.

Combined results from somatic deletion and genetic linkage analyses. When we combined the results from our somatic deletion study and germline linkage study, the results overlapped, implicating two consensus regions at 8p (Fig. 3). One was at 8p21.3 between 20.6 and 23.7 Mb, and the other was at 8p23.1 between 9.8 and 11.2 Mb. Many known and predicted genes are located within these two consensus regions. Five known genes and seven predicted genes are located within the 8p23.1 consensus region. A far greater number of genes are located at the 8p21.3 consensus region, with at least 37 known protein-coding genes. Some of these genes, including *NKX3.1*, have been previously associated with HPC (15).

Interestingly, the 8p21.3 homozygous deletion (between 22.4 and 23.0 Mb) identified in the tumor G6-002 falls within the ~3-Mb 8p21.3 consensus region. Ten known genes are located within this homozygous deletion region.

Discussion

Chromosome 8p has received a great deal of attention from cancer researchers in the past decade because it is commonly deleted in prostate cancer (1, 16) as well as in many other cancers, including colon, breast, ovarian, liver, lung, bladder, and head and neck cancer. Furthermore, results from multiple genetic linkage studies provide evidence that 8p may harbor major prostate cancer susceptibility genes. Considerable efforts have been devoted to the identification of specific prostate cancer genes at 8p that account for the observations from deletion and linkage studies. Although several candidate genes at 8p have been reported to be involved in prostate cancer development, including *NKX3.1, N33, MSR1*, and *DLC1*, few are consistently implicated among different studies. One of the major difficulties is the broad genomic regions implicated in these deletion and linkage studies; for example, most of the observed deletions involve the entire 8p arm. In this study, we used two complementary methodologies in an attempt to effectively narrow the genomic region(s) harboring prostate cancer genes. We used high-resolution Affymetrix SNP arrays to define detailed deletion patterns at 8p among 55 prostate cancers. This analysis led to the identification of two small deleted regions at 8p21.3 and 8p23.1. We did a fine mapping linkage analysis at 8p among 206 HPC families and obtained evidence for linkage at these two regions. Most importantly, we combined the results from the somatic deletion analysis and genetic linkage analysis to further narrow the regions to ~3.1 Mb at 8p21.3 and ~1.4 Mb at 8p23.1. These much smaller consensus regions will likely facilitate more effective searches for prostate cancer genes at 8p.

The high-resolution SNP arrays provide a better tool to identify small DNA copy number alterations and to examine detailed patterns of deletions. With a denser resolution of SNPs covering the 8p region, combined with allele specific analysis, we were able to detect small deletions and better define the boundaries of deletions. In addition, the high density of SNPs revealed interstitial 8p deletions in three cancers. These findings allowed us to identify two small overlapping deleted regions at 8p21.3 and 8p23.1. It is interesting to note that these two separate deleted regions are within the single 27.1-Mb deleted region at 8p identified from a combined analysis of 891 prostate cancers as the most deleted region in the genome (2). Our current study provides evidence that the previously known commonly deleted region may consist of two separate deleted regions.

The fine mapping panel of 29 markers at 8p in our linkage study, with an average of ~1-cM resolution, provides a better tool to dissect detailed linkage patterns among the 206 HPC families. In this study, we were able to confirm prostate cancer linkage at 8p among a large number of prostate cancer families. More importantly, we were able to obtain statistical evidence for two separate linkage regions. One of the linkage regions (19.6–23.9 Mb at 8p21.3) overlapped with the 1-LOD drop interval at 8p21 (22–32 Mb) reported from 1,233 ICPCG prostate cancer families (5).

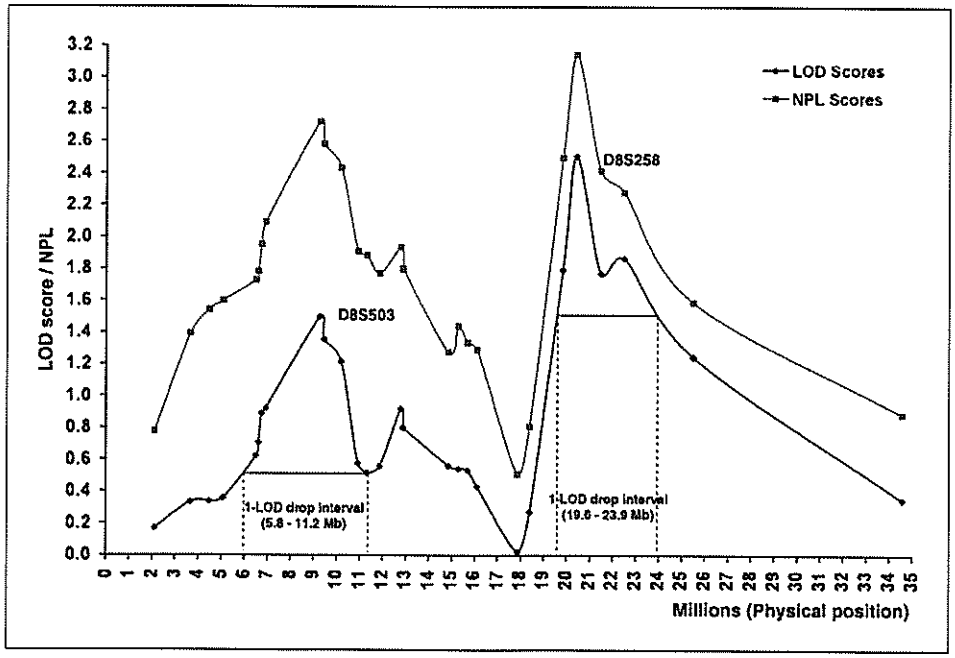

Figure 2. Linkage analysis results at 8p among 206 prostate cancer families. Two primary linkage peaks that we observed. *Y-axis*, LOD (*light purple lines*) and NPL (*dark blue lines*) scores; *X-axis*, physical position along the 8p arm. For the peaks observed at 8p21.3 and 8p23.1, vertical dashed lines have been used to indicate the span of 1-LOD drop intervals.

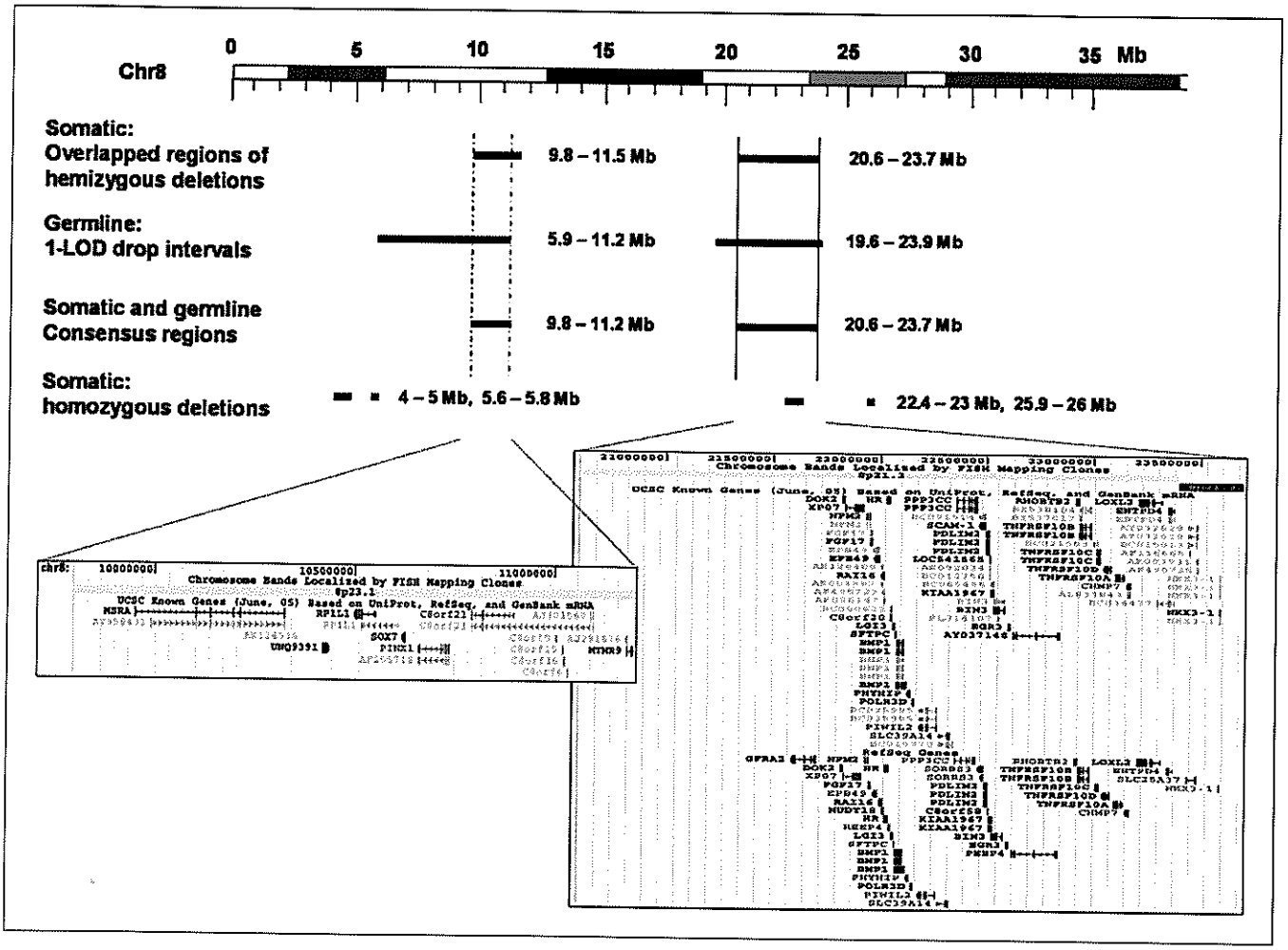

Figure 3. Combined results of somatic deletion study and germline linkage study. Overlapping results provided by our analyses of germline linkage and somatic deletion. Two consensus regions at 8p21.3 and 8p23.1 (*solid vertical lines*). *Bottom,* known and predicted genes located within these two consensus regions, based on release hg18 of the UCSC database.

As hypothesized in the "two-hit" model (10, 11), inherited genetic defects, combined with acquired somatic changes, ultimately alter the expression and/or function of tumor suppressor genes and lead to cancer. Therefore, approaches that combine information from germline and somatic studies may provide better power to identify cancer genes. This combined approach has been successfully used to identify the *APC* gene for familial adenomatous polyposis (FAP). Results from genetic linkage studies in FAP families, somatic loss of heterogeneity analysis, and an interstitial germline deletion all converged to a small region at 5q21 and led to the identification of the *APC* gene (17). Although there are large differences between the rare syndrome of FAP and prostate cancer, the principle of the "two-hit" model may still apply, and our combined approach represents a critical step toward the identification of prostate cancer genes at 8p.

It is interesting that both somatic deletion analysis of prostate tumors and germline linkage analysis of prostate cancer families identified the same genomic regions. Although by no means conclusive, this overlap is consistent with the hypothesis that the same gene or genes is affected both at the germline and somatic levels. Unfortunately, because tumor tissue is not available from the families linked to this region, we can not determine whether the non-linked allele is more likely to undergo somatic deletion, as is observed in multiple inherited cancer syndromes. The observation that most tumors with 8p deletions have deleted both of the implicated regions, and that at least some prostate cancer families are linked to both regions, suggests that multiple genes in these intervals may need to be affected before prostate carcinogenesis can proceed effectively. In any event, the results of this integrated analysis improve the confidence that these two regions most likely contain prostate cancer genes as well as provide more detailed positional information regarding the genomic regions harboring these genes.

In summary, we have combined genetic linkage information with somatic deletion mapping in an attempt to refine the localization of prostate cancer genes on the short arm of chromosome 8. The genomic intervals narrowed by this combined approach provide novel positional information useful for the eventual identification of specific genes important in prostate carcinogenesis.

Acknowledgments

Received 12/12/2005; revised 1/24/2007; accepted 2/9/2007.

Grant support: National Cancer Institute grants CA106523 and CA95052 (J. Xu) and Department of Defense grant PC051264 (J. Xu).

The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked *advertisement* in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

We thank all the study subjects who participated in this study.

References

1. Dong JT. Chromosomal deletions and tumor suppressor genes in prostate cancer [review]. Cancer Metastasis Rev 2001;20:173–93.
2. Sun J, Liu W, Adams TS, et al. DNA copy number alterations in prostate cancers: a combibed analysis of published CGH studies. Prostate 2007 Mar 6; [Epub ahead of print].
3. Liu W, Chang B, Sauvageot J, et al. Comprehensive assessment of DNA copy number alterations in human prostate cancers using Affymetrix 100K SNP mapping array. Genes Chromosomes Cancer 2006;45:1018–32.
4. Xu J, Zheng SL, Hawkins GA, et al. Linkage and association studies of prostate cancer susceptibility: evidence for linkage at 8p22-23. Am J Hum Genet 2001; 69:341–50.
5. Xu J, Dimitrov L, Chang BL, et al. A combined genomewide linkage scan of 1,233 families for prostate cancer-susceptibility genes conducted by the international consortium for prostate cancer genetics. Am J Hum Genet 2005;77:219–29.
6. Knudson AG, Jr. Mutation and cancer: statistical study of retinoblastoma. Proc Natl Acad Sci U S A 1971;68: 820–3.
7. Bova GS, MacGrogan D, Levy A, Pin SS, Bookstein R, Isaacs WB. Physical mapping of chromosome 8p22 markers and their homozygous deletion in a metastatic prostate cancer. Genomics 1996;35:46–54.
8. Nannya Y, Sanada M, Nakazaki K, et al. A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays. Cancer Res 2005;65:6071–9.
9. Lin M, Wei LJ, Sellers WR, Lieberfarb M, Wong WH, Li C. dChipSNP: significance curve and clustering of SNP-array-based loss-of-heterozygosity data. Bioinformatics 2004;20:1233–40.
10. Kong X, Matise TC. MAP-O-MAT: Internet-based linkage mapping. Bioinformatics 2005;21:557–9.
11. Kruglyak L, Daly MJ, Reeve-Daly MP, Lander ES. Parametric and nonparametric linkage analysis: a unified multipoint approach. Am J Hum Genet 1996; 58:1347–63.
12. Kong A, Cox NJ. Allele-sharing models. LOD scores and accurate linkage tests. Am J Hum Genet 1997;61: 1179–88.
13. Smith JR, Freiji D, Carpten JD, et al. A genome wide search reveals a major susceptibility locus for prostate cancer on chromosome 1. Science 1996;274:1371–4.
14. Whittemore A, Halpern J. A class of tests for linkage using affected pedigree members. Biometrics 1994;50: 118–27.
15. Zheng SL, Ju JH, Chang BL, et al. Germ-line mutation of NKX3.1 cosegregates with hereditary prostate cancer and alters the homeodomain structure and function. Cancer Res 2006;66:69–77.
16. Macoska JA, Trybus TM, Benson PD, et al. Evidence for three tumor suppressor gene loci on chromosome 8p in human prostate cancer. Cancer Res 1995;15:5390–5.
17. Kinzler KW, Nilbert MC, Vogelstein B, et al. Identification of a gene located at chromosome 5q21 that is mutated in colorectal cancers. Science 1991;251: 1366–70.

The Prostate 67:692–700 (2007)

DNA Copy Number Alterations in Prostate Cancers: A Combined Analysis of Published CGH Studies

Jishan Sun,[1] Wennuan Liu,[1] Tamara S. Adams,[1] Jielin Sun,[1] Xingnan Li,[1] Aubrey R. Turner,[1] Baoli Chang,[1] Jin Woo Kim,[1] Siqun Lilly Zheng,[1] William B. Isaacs,[2] and Jianfeng Xu[1]*

[1]*Center for Human Genomics, Wake Forest University School of Medicine, Winston-Salem, North Carolina*
[2]*Department of Urology, Johns Hopkins Medical Institutions, Baltimore, Maryland*

BACKGROUND. Identifying genomic regions that are commonly deleted or gained in neoplastic cells is an important approach to identify tumor suppressor genes and oncogenes. Studies in the last two decades have identified a number of common DNA copy number alterations in prostate cancer. However, because of various sample sizes, diverse tumor types and sources, as well as a variety of detection methods with various sensitivities and resolutions, it is difficult to summarize and fully interpret the overall results.

METHODS. We performed a combined analysis of all published comparative genomic hybridization (CGH) studies of prostate cancer and estimated the frequency of alterations across the genome for all tumors, as well as in advanced and localized tumors separately. A total of 41 studies examining 872 cancers were included in this study.

RESULTS. The frequency of deletions and gains were estimated in all tumors, as well as in advanced and localized tumors. Eight deleted and five gained regions were found in more than 10% of the prostate tumors. An additional six regions were commonly deleted and seven were commonly gained in advanced tumors. While 8p was the most common location of deletion, occurring in about a third of all tumors and about half of advanced tumors, 8q was the most commonly gained region, affecting about a quarter of all tumors and about half of all advanced tumors.

CONCLUSIONS. The large number of tumors examined in this combined analysis provides better estimates of the frequency of specific alterations in the prostate cancer cell genome, and offers important clues for prioritizing efforts to identify tumor suppressor genes and oncogenes in these altered regions. *Prostate 67: 692–700, 2007.* © 2007 Wiley-Liss, Inc.

KEY WORDS: **somatic; gain; deletion**

INTRODUCTION

Prostate cancer is the most common cancer among men in developed countries. It is estimated that 234,460 new cases of prostate cancer will be diagnosed in 2006 in the United States [1]. Significant progress has been made in understanding the etiology of this disease in the past few decades. Family history of prostate cancer, age, and race are three well-established risk factors for the disease [2]. Germline and somatic changes in many genes have been reported to be involved in the development of prostate cancer; however, only a few specific genes have been consistently implicated.

Cancers are thought to arise as a result of alterations in expression of tumor suppressor genes and oncogenes in prostate epithelial cells. Altered gene

Grant sponsor: National Cancer Institute; Grant numbers: CA106523, CA95052; Grant sponsor: Department of Defense; Grant number: PC051264.

*Correspondence to: Dr. Jianfeng Xu, Center for Human Genomics, Medical Center Blvd, Winston-Salem, NC 27157.
E-mail: jxu@wfubmc.edu

Received 19 September 2006; Accepted 9 November 2006
DOI 10.1002/pros.20543
Published online 6 March 2007 in Wiley InterScience (www.interscience.wiley.com).

© 2007 Wiley-Liss, Inc.

WILEY InterScience® DISCOVER SOMETHING GREAT

expression may result from one or a combination of factors, including inherited genetic changes, acquired somatic genetic changes, and epigenetic changes such as methylation and imprinting. There are three major types of somatic genetic changes in DNA sequences, including point mutations affecting single bases, small-size deletions or insertions (INDELs), and large-size deletions or gains (DNA copy number alterations).

Various methods have been used to detect somatic DNA copy number alterations in tumors, including cytogenetic evaluation of chromosomal aberrations, DNA polymorphism analysis for detecting loss of heterozygosity (LOH), and comparative genomic hybridization (CGH) approaches for identifying segmental copy number changes. In CGH analysis, differentially labeled tumor DNA and matched normal DNA are co-hybridized to a metaphase chromosome spread (conventional CGH) or a microarray (array-based CGH), such as cDNA array, oligo-nucleotide array, or bacterial artificial chromosome (BAC) array. The ratio of the tumor versus normal hybridization intensities at specific intervals or probes indicates the relative copy number at the genomic location mapped to that interval.

Studies of DNA copy number alterations in prostate cancers have identified multiple frequently altered regions in the genome which has led to the identification of important prostate tumor suppressors and oncogenes. Some examples include *PTEN* at the 10q23 deleted region [3], *ATBF1* at the 16q22 deleted region [4], *KLF5* at the 13q21 deleted region [5], *AR* at the Xq12 gained region among hormone refractory tumors [6], and *MYC* at the 8q21 gained region [7]. However, it is likely that other tumor suppressor genes and oncogenes exist in these frequently altered regions that contribute to the selection advantage that occurs in tumor cells. Furthermore, no specific genes have been identified in other commonly deleted and gained regions, probably due to a combination of relatively larger sizes of affected regions and/or relatively smaller effects of the genes.

One effective approach to improve the ability to identify genes that drive the selection of altered regions in tumors is to increase the study sample size by combining the results from multiple published studies. Many studies on DNA copy number alterations in prostate cancers have been published over the past 20 years, using various molecular methods. However, the different resolutions of these methods in identifying DNA copy number alterations, various sample sizes, and diverse tumor types and sources (localized/primary tumors, metastatic/recurrent tumors, xenografts and cell lines) makes it difficult to fully comprehend and interpret the overall results. Therefore a systematic and uniform approach was needed. In this study, we performed a combined analysis of all published CGH studies of prostate tumors with the intent of estimating the frequencies of DNA copy number alterations in the genome and narrowing these regions to facilitate the identification of genes driving the selection of these alterations.

METHODS

We searched the PubMed database for all published papers on DNA copy number alterations of prostate tumors using CGH methods as of April of 2006. A total of 289 papers were retrieved using the key words "prostate" and "comparative genomic hybridization." The relevance of these papers was determined by reviewing the abstracts, methods, and results sections of each paper. Papers were excluded if the DNA copy number alterations of individual tumor samples were not presented or if they could not be inferred from the text, tables, or graphs. To reduce the chance of the same study samples being counted more than once, the sources of material for each study were carefully examined and then any duplicate samples were removed. In addition, we added four papers from the references cited in these papers. A total of 41 studies and 872 tumors were included in our combined analysis [8–48].

One of the greatest challenges of this combined analysis was to find a uniform approach to address the different resolutions of cytogenetic bands used to report DNA copy number alterations due to various CGH methods in these published studies. To reconcile these differences, we chose to estimate the frequency of DNA copy number alterations based on the 850-band cytogenetic map. The frequency of deletion or gain at each of the 850 bands was estimated based on the number of tumors having deletions or gains among all the tumors examined at each band, respectively. In studies where the results were only graphically reported, we inferred the altered regions from the graphs. The peak of each altered region was defined as the cytogenetic band that has the highest percentage of alterations. The interval of each altered region was defined as regions of cytogenetic bands immediately surrounding the peak with a percentage that was higher than the lower bound of the 95% confidence interval (CI) of the peak.

The frequency of DNA copy number alterations was first estimated in all prostate cancers, including localized/primary tumors, metastatic/recurrent tumors, and cell lines or xenografts. These frequencies were further estimated separately in localized tumors only, including localized/primary tumors (N = 659) and in advanced tumors only, including metastatic/recurrent tumors and prostate cancer cell lines and

xenografts (N = 213). The difference in the frequency of alterations between the two groups at each affected region was tested using two-sample proportion test.

The number of transcripts residing in each implicated region was obtained from the Sanger Institute: http://www.ensembl.org/Multi/martview/9mcHEx6SWc.mart. We searched for cancer-related genes in several databases: (1) Sanger Institute (http://www.sanger.ac.uk/genetics/CGP/Census/chromosome.shtml), (2) Atlas of genetics and Cytogenetics in Oncology and Heamatology (http://atlasgeneticsoncology.org), and (3) ExPASy Proteomics Server (http://us.expasy.org). We also included additional candidate cancer-related genes in the implicated regions based on published reports.

RESULTS

In total, 872 independent prostate tumors, including 51 cell lines and xenografts, were analyzed for DNA copy number alterations using CGH methods in these 41 papers. Most of these tumors were examined for DNA copy number alterations in the entire genome; however, several tumors were examined for alterations in specific chromosomes or chromosomal arms. Therefore, the number of tumors examined at specific cytogenetic bands was different across the genome.

The frequency of deletions and gains across the genome in all tumors, as well as in advanced and localized tumors were estimated (Fig. 1a,b, and supplement Tables I and II). There were 13 regions in the prostate tumor genome that were frequently altered, defined as those observed in more than 10% of the examined tumors. Among these common alterations, eight were deletions and five were gains. In addition, six additional regions, including three deletions and three gains, were found in more than 10% of advanced tumors. The peak and interval for each of these altered regions are presented in Table I.

Chromosome 8p was the most commonly deleted region in the genome. The peak was observed at 8p21.3 (19.1–23.4 Mb); 284 of 833 (34.09%) examined tumors had a deletion at this cytogenetic band. The deletion pattern was unimodal; the frequency of deletion dropped slightly at either side of the peak. Seven cytogenetic bands (8p23.1–8p21.1, from 6.2 to 28.9 Mb) immediately surrounding the peak had a frequency of deletion that was higher than the lower bound of the 95% CI of estimate at the peak (30.87%). In addition to this defined interval, the deletion curve extended gradually at the centromeric side for five additional cytogenetic bands (percentage slightly dropped to ~17.5% at 8p11.1). The curve dropped sharply to below 3% beyond the centromere. The second most commonly deleted region was observed at 13q21.31 (61.2–64.6 Mb); 228 of 813 (28.04%) examined tumors had a deletion at this cytogenetic band. The deletion pattern was also unimodal and broad; nine cytogenetic bands (13q14.13–13p22.1, from 44.7 to 74.2 Mb) immediately surrounding the peak had an estimated frequency of deletion higher than the lower bound of the 95% CI of the estimate at the peak (24.96%). Other commonly deleted regions were at 6q14.1–6q21 (22.24%), 16q13–16q24.3 (17.85%), 18q12.1–18q23 (12.80%), 5q13.3–5q21.3 (13.06%), 2q21.2–2q22.3 (12.42%), and 10q23.1–10q25.3 (11.76%).

Chromosome 8q was the most commonly gained region in the genome of prostate tumors. The pattern of gain at this region was bimodal. The first peak was observed at 8q22.2 (99.1–101.6 Mb); 215 of 857 (25.09%) examined tumors had a gain at this cytogenetic band. The second peak was on the telomeric side at 8q24.13 (122.5–124.9 Mb); 201 of 837 (24.01%) examined tumors had a gain at this cytogenetic band. The interval of the 8q gain was broad; 12 cytogenetic bands (8q21.3–8q24.3, from 87 to 143.1 Mb) immediately surrounding the peak had frequency of gain higher than the lower bound of 95% CI of estimate at the peak (22.18%). The frequencies of the remaining commonly gained regions were all considerably lower (7%~12%). They were 7q11.21–7q32.3 (12.48%), Xq11.1–Xq23 (10.86%), 17q24.1–17q25.3 (11.65%), and 3q23–3q26.33 (10.24%).

Similar findings were observed when the combined analyses were performed separately for advanced and localized tumors. All of the commonly deleted and gained regions identified in all tumor samples were frequently altered in both advanced and localized tumors. The frequencies of these altered regions were however two- to threefold higher in advanced tumors than localized tumors (Fig. 1a,b). The differences in the frequencies between these two groups were statistically significant; the Z-scores ranged from 3.00 to 11.64 (P values 0.003–10^{-13}) for the eight deletion regions and from 2.96 to 7.06 (P values 0.003 to ~0) for the five gained regions. For example, at the 8p deleted region, 56.25% of the advanced tumors were deleted at 8p21.3, compared with 28.00% of the localized tumors that were deleted, Z-score = 7.40, P-value = 1.7×10^{-13}. At the 8q gain region, 47.87% of the advanced tumors had gains at 8q22.1, compared with 17.10% of the localized tumors having gains at this cytogenetic band, Z-score = 8.69, P-value ~0.

In addition to the regions that were implicated in all tumors, there were several novel altered regions implicated in advanced tumors only. Six additional regions were deleted in more than 10% of advanced tumors, including 15q15.1–15q23 (21.93%), 4q13.1–4qter (22.30%), 22q11.1–22qter (15.48%), 1pter–1p21.1 (14.89%), 9q21.13–9qter (14.19%), and 12p (12.24%).

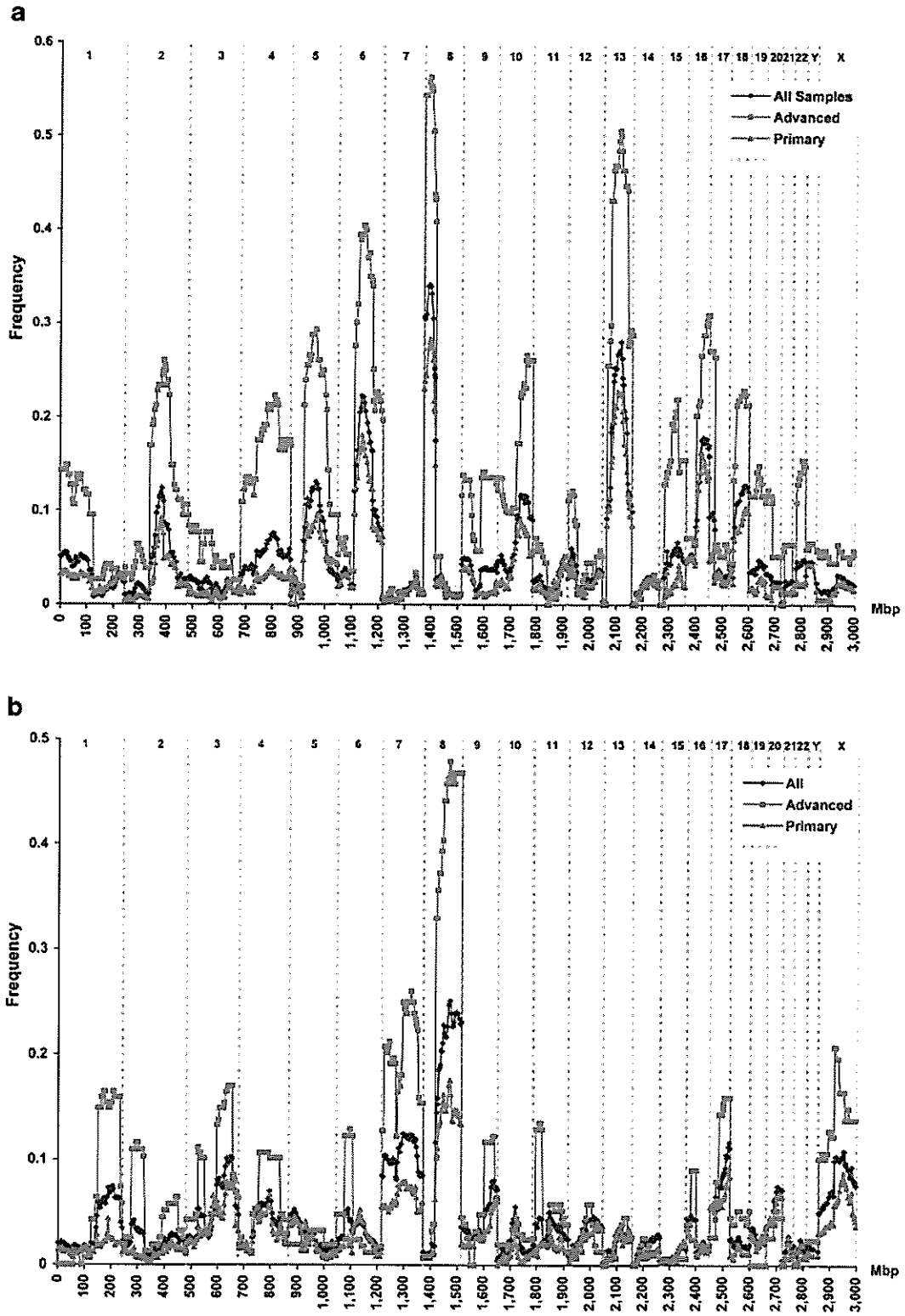

Fig. 1. The combined frequencies of DNA copy number alterations in the genome of prostate cancers. The X-axis represents the position of the genome in Mb, and theY-axis represents the frequency of alterations for deletions (**a**) and gains (**b**). Each chromosome was designated by its corresponding number and the divisions between individual chromosomes are shown by vertical lines. Frequencies in all tumors (N = 872), localized tumors (N = 659), and advanced tumors and cell lines (N = 225),including metastatic/recurrent tumors (N = 174) as well as prostate cancer cell lines and xenografts (N = 51), were plotted separately as represented by diamonds, squares, and triangles, respectively. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

The Prostate DOI 10.1002/pros

TABLE I. Characterizations of the Frequently Altered Chromosomal Regions in the Prostate Cancers

Cytogenetic bands							
Peak	Start	End	Interval length (Mb)	Peak frequency (%)	# of coding[b]	Tumor related genes protein[c]	References
(Loss)							
2q22.2	2q21.2	2q22.3	16.2	12.42	41	*TTL, ERCC3, BIN1*	[9,11,13,14,16–18,20,22–24,27–29,33,34,37–39,41,43–47]
4q27–4q28.1[a]	4q21.3	4q31.1	59.5	7.55	208	*MAD2L1, SYNPO2, MAPK10*	[9,13,16,20,23,25,27–29,34,36,38,39,41,43–47]
5q15	5q13.3	5q21.3	31.9	13.06	142	*APC, MCC*	[9,14,16–18,20,22–25,27–29,34,36,39,41–43,45–48]
6q15	6q14.1	6q21	29.9	22.24	170	*FOXO3A, GOPC, MAP3K7**	[10,13,14,16–20,22–25,27–29,33,34,36–44,46–48]
8p21.3	8p23.1	8p21.1	22.7	34.09	238	*WRN, WHSC1L1, FGFR1, LZTS1, TUSC3*, MSR1*, NKX3.1*, DLC1**	[8,9,11–18,20–29,32–37,39–43,45–48]
10q23.2	10q23.1	10q25.3	31.9	11.76	287	*PTEN, TNFRSF6, SUFU, FGFR2*	[9,13–17,20,23,24,26,27,29,30,33,34,39,41–43,46,48]
12p13.1[a]	12p13.32	12p11.23	24.6	6.02	253	CDKN1B, RECQL, CCND2	[11,16,20,23,27,34,44,46]
13q21.33	13q14.13	13q22.1	29.5	28.04	95	*RB1, FOXO1A, LCP1, RFP2, DLEU1, DLEU2, KLF5*	[9,11,13,14,16–29,33–36,38–48]
15q23[a]	15q21.1	15q25.3	44.3	6.66	398	*TCF12, BUB1B, THBS1*	[9,13,14,16,17,20,27,29,33,34,39,41,43,45]
16q22.1	16q13	16q24.3	32	17.85	346	*ERCC4, CYLD, CBFB, CDH1, ATBF1, CDH11, CBFA2T3, FANCA, RBL2*	[9,11,13,14,16–18,20–25,27,29,30,33,34,36,37,39–43,46–48]
18q21.33–22.1	18q12.1	18q23	46.5	12.80	200	*SERPINB5, SMAC, SMAD4, MALT1, DCC*	[9,11,13,14,16–18,20–29,34,36,37,40,41,43,44,46–48]
(Gain)							
1q32.2[a]	1q24.1	1q42.3	70.6	7.46	530	*SIL, ABL2*	[9,16–18,20,28,29,34,36,39,40,43,45–47]
3q26.1	3q23	3q26.33	43.8	10.24	213	*GMPS, PIK3CA, MLF1, SKIL, CCNL1, ECT2*	[9,11,13,14,16–18,20,21,23,27–29,34,36–39,42,43,47]
4q26[a]	4q21.21	4q27	45.1	6.99	208	*NFKB1*	[11,13,17,23,30,34,35,39,40,42,43,47,48]
7q21.11	7q11.21	7q33	75.9	12.48	535	*CDK6, SMOH, TIF1, HIP1, ABCB1, MET*	[8,9,11,13,14,16–18,22,23,27–29,32–34,36–41,43,45–48]
8q22.2	8q21.3	8q24.3	52.9	25.09	303	*CBFA2T1, WISP1, MYC*	[9–11,13,14,16–18,20,21,23,24,26,27,29,30,32–37,39–43,45–48]
9q33.3[a]	9q32	9q34.3	26.4	8.07	382	*ABL1, SET*	[9,11,13,14,16–18,20,23,24,26–29,39,40,42,43,45–48]
17q25.2	17q24.1	17q25.3	15.0	11.65	281	*MAFG, BCAS3*	[9,11,13,14,16–18,20,22–25,27–29,34,39,40,43,45–48]
Xq21.33	Xq11.1	Xq23	56.9	10.86	282	*AR**	[13,16–18,27,29,30,33,34,36,37,40,41,44–47]

[a]Regions denoted by 'a' represent the altered copy number frequencies are below 10% in all tumors but above 10% in advanced tumors.
[b]The numbers of coding proteins were obtained from http://www.ensembl.org/Multi/martview/frKYjCwJyF.mart.
[c]Cancer related genes, with exception for those that are underlined or marked with asterisks, were defined by three online databases described in the text. Genes that were underlined are not located in but near the defined regions. Genes with asterisks are not included in the three databases, but their relevance to prostate cancer was proposed in the literature.

Seven additional regions had gains in more than 10% of the advanced tumors, including 11p15. 3–11p13 (13.55%), 1q21.3–1q42.3 (16.49%), 2p22.3–2p12 (11.61%), 9q22.11–9q33.3 (12.23%), 3p21.33–3p14.1 (11.17%), 4q21.1–4q31.3 (10.64%), and 6p21.33–6p11.1 (12.92%).

We searched three public databases for tumor related genes in these commonly implicated regions. We also included additional candidate cancer-related genes that were not included in the databases but were implicated in published literature. These results are presented in Table I.

DISCUSSION

In an attempt to provide a systematic summary of previously identified DNA copy number alterations in prostate cancer, we performed a combined analysis of all CGH studies of prostate cancer available on PubMed. The large number of cancers examined in this combined study and the use of a unified cytogenetic map should improve the ability to accurately estimate the frequencies of DNA copy number alterations in the genome.

In this analysis, we found eight commonly deleted regions and five commonly gained regions in prostate tumors. We also found six additional deleted regions and seven gained regions that are commonly implicated in advanced tumors. Our findings on common deletions are generally consistent with the results of a previous review paper [49] where the 8p deletion was found in 175 of 417 tumors (42%). To our knowledge, our finding on the commonly gained regions across the genome of prostate tumors is the first of its kind in the published literature. In addition, our study is the first to estimate the percentage of deletions and gains present in advanced and localized tumors.

The results of our combined analysis indicate that 8p is the most commonly deleted region in the prostate tumor genome, affecting about a third of all tumors and half of advanced tumors. The second and third most commonly deleted regions are 13q and 6q, slightly less frequent than 8p. On the other hand, our results clearly show that 8q is the most commonly gained region, affecting about a quarter of all tumors and half of advanced tumors. The frequency of 8q gain is considerably higher than that of 7q, the second most commonly gained region which affects about 10% of all tumors and nearly a quarter of advanced tumors. The estimated percentage of alterations at each commonly implicated region and their rank in the genome should provide important clues in prioritizing our efforts to identifying tumor suppressor genes and oncogenes in these altered regions.

Cancer related genes in the intervals of the commonly altered regions, as shown in Table I, are reasonable candidate tumor suppressor genes and oncogenes, and therefore they warrant further examination. For example, loss of the *PTEN* gene copy at 10q24 commonly reflects a loss of PTEN tumor suppressive function and drives the selection of cells that have alterations at this region [3]. Similarly, the gain of *AR* gene copies at the Xq12 commonly gained region may lead to over expression of AR and may be responsible for prostate tumor progression, particularly towards androgen independent disease. In addition to the known candidate genes, it is quite likely that multiple novel prostate tumor suppressors and oncogenes, located in regions of recurrent copy number alteration, remain to be identified.

It is important to note the limitations of our study. As a combined analysis of published papers, our study is subject to publication biases. It is possible that studies are more likely to be written up and accepted for publication if they found results similar to those of previously published studies, and this may inflate our estimated frequencies of the most commonly implicated regions. On the other hand, results on novel altered regions may be under-represented in publications. This may explain the low frequency of deletions reported at 21q22, which was recently found to be commonly deleted in prostate cancers [50–52]. In addition, our results are also affected by the limited resolution of the methods used to identify DNA copy number alterations in the original published papers. The vast majority of these published studies used conventional CGH methods; the limited resolution of these methods (~10 Mb),may affect the ability to detect small size alterations. Furthermore, the limited resolution may also contribute to the relatively poor ability to narrow the altered regions in our study. A small number of tumors (<8%) in this study were analyzed using the higher resolution approach of array-CGH methods [8–10,12,14,15,18,53–55]. While these had better ability to detect small size deletions, the sample size was too small to warrant a separate analysis and to influence the overall results.

Future studies of DNA copy number alterations should use methods with higher resolution. Several high-resolution genome-wide analyses have recently been developed for assessing DNA copy number alterations, including representational oligonucleotide microarray analysis (ROMA) and Affymetrix SNP mapping arrays [50,56–60]. These newer molecular methods will improve our ability to accurately identify DNA copy number alterations in prostate tumors, especially smaller alterations, and may eventually lead to the discovery of prostate tumor suppressor genes and oncogenes.

ACKNOWLEDGMENTS

The study is partially supported by National Cancer Institute CA106523 (to J.X.), CA95052 (to J.X.), and Department of Defense grant PC051264 (to J.X.).

REFERENCES

1. Jemal A, Siegel R, Ward E, Murray T, Xu J, Smigal C, Thun MJ. Cancer statistics. CA Cancer J Clin 2006;56(2):106–130.
2. Isaacs WB, Xu J. Prostate cancer. In: King RA, Totter JI, Motulsky AG, editors. The genetic basis of common diseases. New York: Oxford University Press; 2002;738–748.
3. Li J, Yen C, Liaw D, Podsypanina K, Bose S, Wang SI, Puc J, Miliaresis C, Rodgers L, McCombie R, Bigner SH, Giovanella BC, Ittmann M, Tycko B, Hibshoosh H, Wigler MH, Parsons R. PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. Science 1997; 275(5308):1943–1947.
4. Sun X, Frierson HF, Chen C, Li C, Ran Q, Otto KB, Cantarel BL, Vessella RL, Gao AC, Petros J, Miura Y, Simons JW, Dong JT. Frequent somatic mutations of the transcription factor ATBF1 in human prostate cancer. Nat Genet 2005;37(4):407–412.
5. Chen C, Bhalala HV, Vessella RL, Dong JT. KLF5 is frequently deleted and down-regulated but rarely mutated in prostate cancer. Prostate 2003;55(2):81–88.
6. Visakorpi T, Hyytinen E, Koivisto P, Tanner M, Keinanen R, Palmberg C, Palotie A, Tammela T, Isola J, Kallioniemi OP. In vivo amplification of the androgen receptor gene and progression of human prostate cancer. Nat Genet 1995;9(4):401–4066.
7. Jenkins RB, Qian J, Lieber MM, Bostwick DG. Detection of c-myc oncogene amplification and chromosomal anomalies in metastatic prostatic carcinoma by fluorescence in situ hybridization. Cancer Res 1997;57(3):524–531.
8. Brookman-Amissah N, Duchesnes C, Williamson MP, Wang Q, Ahmed A, Feneley MR, Mackay A, Freeman A, Fenwick K, Iravani M, Weber B, Ashworth A, Masters JR. Genome-wide screening for genetic changes in a matched pair of benign and prostate cancer cell lines using array CGH. Prostate Cancer Prostatic Dis 2005;8(4):335–343.
9. Paris PL, Albertson DG, Alers JC, Andaya A, Carroll P, Fridlyand J, Jain AN, Kamkar S, Kowbel D, Krijtenburg PJ, Pinkel D, Schroder FH, Vissers KJ, Watson VJ. High-resolution analysis of paraffin-embedded and formalin-fixed prostate tumors using comparative genomic hybridization to genomic microarrays. Am J Pathol 2003;162(3):763–770.
10. van Duin M, van Marion R, Watson JE, Paris PL, Lapuk A, Brown N, Oseroff VV, Albertson DG, Pinkel D, de Jong P, Nacheva EP, Dinjens W, van Dekken H, Collins C. Construction and application of a full-coverage, high-resolution, human chromosome 8q genomic microarray for comparative genomic hybridization. Cytometry A 2005;63(1):10–19.
11. Teixeira MR, Ribeiro FR, Eknaes M, Waehre H, Stenwig AE, Giercksky KE, Heim S, Lothe RA. Genomic analysis of prostate carcinoma specimens obtained via ultrasound-guided needle biopsy may be of use in preoperative decision-making. Cancer 2004;101(8):1786–1793.
12. Yano S, Matsuyama H, Matsuda K, Matsumoto H, Yoshihiro S, Naito K. Accuracy of an array comparative genomic hybridization (CGH) technique in detecting DNA copy number aberrations: Comparison with conventional CGH and loss of heterozygosity analysis in prostate cancer. Cancer Genet Cytogenet 2004;150(2):122–127.
13. Strohmeyer DM, Berger AP, Moore DH 2nd, Bartsch G, Klocker H, Carroll PR, Loening SA, Jensen RH. Genetic aberrations in prostate carcinoma detected by comparative genomic hybridization and microsatellite analysis: Association with progression and angiogenesis. Prostate 2004;59(1):43–58.
14. van Dekken H, Paris PL, Albertson DG, Alers JC, Andaya A, Kowbel D, van der Kwast TH, Pinkel D, Schroder FH, Vissers KJ, Wildhagen MF, Collins C. Evaluation of genetic patterns in different tumor areas of intermediate-grade prostatic adenocarcinomas by high-resolution genomic array analysis. Genes Chromosomes Cancer 2004;39(3):249–256.
15. Hermans KG, van Alewijk DC, Veltman JA, van Weerden W, van Kessel AG, Trapman J. Loss of a small region around the PTEN locus is a major chromosome 10 alteration in prostate cancer xenografts and cell lines. Genes Chromosomes Cancer 2004; 39(3):171–184.
16. Laitinen S, Karhu R, Sawyers CL, Vessella RL, Visakorpi T. Chromosomal aberrations in prostate cancer xenografts detected by comparative genomic hybridization. Genes Chromosomes Cancer 2002;35(1):66–73.
17. Chu LW, Troncoso P, Johnston DA, Liang JC. Genetic markers useful for distinguishing between organ-confined and locally advanced prostate cancer. Genes Chromosomes Cancer 2003; 36(3):303–312.
18. Pettus JA, Cowley BC, Maxwell T, Milash B, Stephenson RA, Rohr LR, Hoff C, Brothman AR. Multiple abnormalities detected by dye reversal genomic microarrays in prostate cancer: A much greater sensitivity than conventional cytogenetics. Cancer Genet Cytogenet 2004;154(2):110–118.
19. Verhagen PC, Hermans KG, Brok MO, van Weerden WM, Tilanus MG, de Weger RA, Boon TA, Trapman J. Deletion of chromosomal region 6q14-16 in prostate cancer. Int J Cancer 2002;102(2):142–147.
20. Kasahara K, Taguchi T, Yamasaki I, Kamada M, Yuri K, Shuin T. Detection of genetic alterations in advanced prostate cancer by comparative genomic hybridization. Cancer Genet Cytogenet 2002;137(1):59–63.
21. Beheshti B, Vukovic B, Marrano P, Squire JA, Park PC. Resolution of genotypic heterogeneity in prostate tumors using polymerase chain reaction and comparative genomic hybridization on microdissected carcinoma and prostatic intraepithelial neoplasia foci. Cancer Genet Cytogenet 2002;137(1): 15–22.
22. Wolter H, Trijic D, Gottfried HW, Mattfeldt T. Chromosomal changes in incidental prostatic carcinomas detected by comparative genomic hybridization. Eur Urol 2002;41(3):328–334.
23. Steiner T, Junker K, Burkhardt F, Braunsdorf A, Janitzky V, Schubert J. Gain in chromosome 8q correlates with early progression in hormonal treated prostate cancer. Eur Urol 2002;41(2):167–171.
24. Wolter H, Gottfried HW, Mattfeldt T. Genetic changes in stage pT2N0 prostate cancer studied by comparative genomic hybridization. BJU Int 2002;89(3):310–316.
25. Verdorfer I, Hobisch A, Culig Z, Hittmair A, Bartsch G, Erdel M, Duba HC, Utermann G. Combined study of prostatic carcinoma by classical cytogenetic analysis and comparative genomic hybridization. Int J Oncol 2001;19(6):1263–1270.
26. Chu LW, Pettaway CA, Liang JC. Genetic abnormalities specifically associated with varying metastatic potential of prostate cancer cell lines as detected by comparative genomic hybridization. Cancer Genet Cytogenet 2001;127(2):161–167.
27. El Gedaily A, Bubendorf L, Willi N, Fu W, Richter J, Moch H, Mihatsch MJ, Sauter G, Gasser TC. Discovery of new DNA

amplification loci in prostate cancer by comparative genomic hybridization. Prostate 2001;46(3):184–190.

28. Zitzelsberger H, Engert D, Walch A, Kulka U, Aubele M, Hofler H, Bauchinger M, Werner M. Chromosomal changers during development and progression of prostate adenocarcinoma. British J Cancer 2001;84(2):202–208.
29. Pan Y, Lui WO, Nupponen N, Larsson C, Isola J, Visakorpi T, Bergerheim US, Kytola S. 5q11, 8p11, and 10q22 are recurrent chromosomal breakpoints in prostate cancer cell lines. Genes Chromosomes Cancer 2001;30(2):187–195.
30. Verhagen PC, Zhu XL, Rohr LR, Cannon-Albright LA, Tavtigian SV, Skolnick MH, Brothman AR. Microdissection, DOP-PCR, and comparative genomic hybridization of paraffin-embedded familial prostate cancers. Cancer Genet Cytogenet 2000;122(1): 43–48.
31. Sattler HP, Lensch R, Rohde V, Zimmer E, Meese E, Bonkhoff H, Retz M, Zwergel T, Bex A, Stoeckle M, Wullich B. Novel amplification unit at chromosome 3q25-q27 in human prostate cancer. Prostate 2000;45(3):207–215.
32. Van Dekken H, Krijtenburg PJ, Alers JC. DNA in situ hybridization (interphase cytogenetics) versus comparative genomic hybridization (CGH) in human cancer: Detection of numerical and structural chromosome aberrations. Acta Histochem 2000; 102(1):85–94.
33. Sattler HP, Rohde V, Bonkhoff H, Zwergel T, Wullich B. Comparative genomic hybridization reveals DNA copy number gains to frequently occur in human prostate cancer. Prostate 1999;39(2):79–86.
34. Nupponen NN, Kakkola L, Koivisto P, Visakorpi T. Genetic alterations in hormone-refractory recurrent prostate carcinomas. Am J Pathol 1998;153(1):141–148.
35. Cher ML, Lewis PE, Banerjee M, Hurley PM, Sakr W, Grignon DJ, Powell IJ. A similar pattern of chromosomal alterations in prostate cancers from African-Americans and Caucasian Americans. Clin Cancer Res 1998;4(5):1273–1278.
36. Nupponen NN, Hyytinen ER, Kallioniemi AH, Visakorpi T. Genetic alterations in prostate cancer cell lines detected by comparative genomic hybridization. Cancer Genet Cytogenet 1998;101(1):53–57.
37. Williams BJ, Jones E, Kozlowski JM, Vessella R, Brothman AR. Comparative genomic hybridization and molecular cytogenetic characterization of two prostate cancer xenografts. Genes Chromosomes Cancer 1997;18(4):299–304.
38. Hyytinen ER, Thalmann GN, Zhau HE, Karhu R, Kallioniemi OP, Chung LW, Visakorpi T. Genetic changes associated with the acquisition of androgen-independent growth, tumorigenicity and metastatic potential in a prostate cancer model. Br J Cancer 1997;75(2):190–195.
39. Cher ML, Bova GS, Moore DH, Small EJ, Carroll PR, Pin SS, Epstein JI, Isaacs WB, Jensen RH. Genetic alterations in untreated metastases and androgen-independent prostate cancer detected by comparative genomic hybridization and allelotyping. Cancer Res 1996;56(13):3091–3102.
40. Joos S, Bergerheim US, Pan Y, Matsuyama H, Bentz M, du Manoir S, Lichter P. Mapping of chromosomal gains and losses in prostate cancer by comparative genomic hybridization. Genes Chromosomes Cancer 1995;14(4):267–276.
41. Visakorpi T, Kallioniemi AH, Syvanen AC, Hyytinen ER, Karhu R, Tammela T, Isola JJ, Kallioniemi OP. Genetic changes in primary and recurrent prostate cancer by comparative genomic hybridization. Cancer Res 1995;55(2):342–347.
42. Cher ML, MacGrogan D, Bookstein R, Brown JA, Jenkins RB, Jensen RH. Comparative genomic hybridization, allelic imbalance, and fluorescence in situ hybridization on chromosome 8 in prostate cancer. Genes Chromosomes Cancer 1994;11(3):153–162.
43. Alers JC, Rochat J, Krijtenburg PJ, Hop WC, Kranse R, Rosenberg C, Tanke HJ, Schroder FH, van Dekken H. Identification of genetic markers for prostatic cancer progression. Lab Invest 2000;80(6):931–942.
44. Koivisto PA, Schleutker J, Helin H, Ehren-van Eekelen C, Kallioniemi OP, Trapman J. Androgen receptor gene alterations and chromosomal gains and losses in prostate carcinomas appearing during finasteride treatment for benign prostatic hyperplasia. Clin Cancer Res 1999;5(11):3578–3582.
45. Zhao H, Kim Y, Wang P, Lapointe J, Tibshirani R, Pollack JR, Brooks JD. Genome-wide characterization of gene expression variations and DNA copy number changes in prostate cancer cell lines. Prostate 2005;63(2):187–197.
46. Fu W, Bubendorf L, Willi N, Moch H, Mihatsch MJ, Sauter G, Gasser TC. Genetic changes in clinically organ-confined prostate cancer by comparative genomic hybridization. Urology 2000; 56(5):880–885.
47. Alers JC, Krijtenburg PJ, Vis AN, Hoedemaeker RF, Wildhagen MF, Hop WC, van Der Kwast TT, Schroder FH, Tanke HJ, van Dekken H. Molecular cytogenetic analysis of prostatic adenocarcinomas from screening studies: Early cancers may contain aggressive genetic features. Am J Pathol 2001;158(2):399–406.
48. Rokman A, Koivisto PA, Matikainen MP, Kuukasjarvi T, Poutiainen M, Helin HJ, Karhu R, Kallioniemi OP, Schleutker J. Genetic changes in familial prostate cancer by comparative genomic hybridization. Prostate 2001;46(3):233–239.
49. Dong JT. Chromosomal deletions and tumor suppressor genes in prostate cancer. Cancer Metastasis Rev 2001;20(3–4):173–193.
50. Liu W, Chang B, Sauvageot J, Dimitrov L, Gielzak M, Li T, Yan G, Sun J, Sun J, Adams TS, Turner AR, Kim JW, Meyers DA, Zheng SL, Isaacs WB, Xu J. Comprehensive assessment of DNA copy number alterations in human prostate cancers using Affymetrix 100 K SNP mapping array. Genes Chromosomes Cancer 2006;45(11):1018–1032.
51. Perner S, Demichelis F, Beroukhim R, Schmidt FH, Mosquera JM, Setlur S, Tchinda J, Tomlins SA, Hofer MD, Pienta KG, Kuefer R, Vessella R, Sun XW, Meyerson M, Lee C, Sellers WR, Chinnaiyan AM, Rubin MA. TMPRSS2:ERG fusion-associated deletions provide insight into the heterogeneity of prostate cancer. Cancer Res 2006;66(17):8337–8341.
52. Yoshimoto M, Joshua AM, Chilton-Macneill S, Bayani J, Selvarajah S, Evans AJ, Zielenska M, Squire JA. Three-color FISH analysis of TMPRSS2/ERG fusions in prostate cancer indicates that genomic microdeletion of chromosome 21 is associated with rearrangement. Neoplasia 2006;8(6):465–469.
53. Saramäki OR, Porkka KP, Vessella RL, Visakorpi T. Genetic aberrations I prostate cancer by microarray analysis. Int J Cancer 2006;119(6):1322–1329.
54. van Duin M, van Marion R, Vissers K, Watson JE, van Weerden WM, Schröder FH, Hop WCJ, van der Kwast TH, Collins C, van Dekken H. High-resolution array comparative genomic hybridization of chromosome arm 8q: Evaluation of genetic progression markers for prostate cancer. Genes Chromosomes Cancer 2005;44(4):438–449.
55. Paris PL, Andaya A, Fridlyand J, Jain AN, Weinberg V, Kowbel D, Brebner H, Simko J, Watson JE, Volik S, Albertson DG, Pinkel D, Alers JC, van der Kwast TH, Vissers KJ, Schroder FH, Wildhagen MF, Febbo PG, Chinnaiyan AM, Pienta KJ, Carroll PR, Rubin MA, Collins C, van Dekken H. Whole genome scanning identifies genotypes associated with recurrence and

metastasis in prostate tumors. Human Mol Genetics 2004;13(13): 1303–1313.

56. Liu W, Chang B, Li T, Dimitrov L, Kim S, Kim JW, Turner AR, Meyers DA, Trent JM, Zheng S, Isaacs WB, Xu J. Germline copy number polymorphisms involving larger than 100 kb are uncommon in normal subjects. Prostate 2007;67(3):227–233.
57. Lucito R, Healy J, Alexander J, Reiner A, Esposito D, Chi M, Rodgers L, Brady A, Sebat J, Troge J, West JA, Rostan S, Nguyen KC, Powers S, Ye KQ, Olshen A, Venkatraman E, Norton L, Wigler M. Representational oligonucleotide microarray analysis: A high-resolution method to detect genome copy number variation. Genome Res 2003;13(10):2291–2305.
58. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, Chi M, Navin N, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A, Wigler M. Large-scale copy number polymorphism in the human genome. Science 2004;305(5683):525–528.
59. Slater HR, Bailey DK, Ren H, Cao M, Bell K, Nasioulas S, Henke R, Choo KH, Kennedy GC. High-resolution identification of chromosomal abnormalities using oligonucleotide arrays containing 116,204 SNPs. Am J Hum Genet 2005;77(5):709–726.
60. Zhao X, Weir BA, LaFramboise T, Lin M, Beroukhim R, Garraway L, Beheshti J, Lee JC, Naoki K, Richards WG, Sugarbaker D, Chen F, Rubin MA, Janne PA, Girard L, Minna J, Christiani D, Li C, Sellers WR, Meyerson M. Homozygous deletions and chromosome amplifications in human lung carcinomas revealed by single nucleotide polymorphism array analysis. Cancer Res 2005;65(13):5561–5570.

GENES, CHROMOSOMES & CANCER 45:1018–1032 (2006)

Comprehensive Assessment of DNA Copy Number Alterations in Human Prostate Cancers Using Affymetrix 100K SNP Mapping Array

Wennuan Liu,[1] Baoli Chang,[1] Jurga Sauvageot,[2] Latchezar Dimitrov,[1] Marta Gielzak,[2] Tao Li,[1] Guifang Yan,[2] Jishan Sun,[1] Jielin Sun,[1] Tamara S. Adams,[1] Aubrey R. Turner,[1] Jin Woo Kim,[1] Deborah A. Meyers,[1] Siqun Lilly Zheng,[1] William B. Isaacs,[2]* and Jianfeng Xu[1,3]*

[1]Center for Human Genomics, Wake Forest University School of Medicine, Winston-Salem, NC
[2]Department of Urology, Johns Hopkins Medical Institutions, Baltimore, MD
[3]Translational Genomics Research Institute (TGen), Phoenix, AZ

Although multiple recurrent chromosomal alterations have been identified in prostate cancer cells, the specific genes driving the apparent selection of these changes remain largely unknown. In part, this uncertainty is due to the limited resolution of the techniques used to detect these alterations. In this study, we applied a high-resolution genome-wide method, Affymetrix 100K SNP mapping array, to screen for somatic DNA copy number (CN) alterations among 22 pairs of samples from primary prostate cancers and matched nonmalignant tissues. We detected 355 recurrent deletions and 223 recurrent gains, many of which were novel. As expected, the sizes of novel alterations tend to be smaller. Importantly, among tumors with increasing grade, Gleason sum 6, 7, and 8, we found a significant trend of larger number of alterations in the tumors with higher grade. Overall, gains are significantly more likely to occur within genes (74%) than are deletions (49%). However, when we looked at the most frequent CN alterations, defined as those in $\geq$4 subjects, we observed that both gains (85%) and deletions (57%) occur preferentially within genes. An example of a novel, recurrent alteration observed in this study was a deletion between the *ERG* and *TMPRSS2* genes on chromosome 21, presumably related to the recently identified fusion transcripts from these two genes. Results from this study provide a basis for a systematic and comprehensive cataloging of CN alterations associated with grades of prostate cancer, and the subsequent identification of specific genes that associated with initiation and progression of the disease. This article contains supplementary material available via the Internet at http://www.interscience.wiley.com/jpages/1045-2257/suppmat © 2006 Wiley-Liss, Inc.

INTRODUCTION

Prostate cancer is the most common cancer among men in the USA. Approximately 235,000 American men may be diagnosed with prostate cancer, corresponding to 33% of all cancer cases, and about 27,000 may die of the disease in 2006 according to the American Cancer Society (Jemal et al., 2006). The lifetime probability of developing prostate cancer for men is one in six in USA, the highest in comparison to other cancers.

Prostate cancer is a heterogeneous collection of subgroups of cancer that display radically different clinical behavior. While some of prostate cancers are capable of dissemination leading to death, some are relatively indolent (Cooperberg et al., 2003). Some are hormone-sensitive. Some are androgen-independent. Some are associated with inherited alterations while others are somatically acquired (Gonzalgo and Isaacs, 2003). The development of these different types of prostate cancers may be regulated by different mechanisms. Correspondingly, different therapeutic targets and strategies should be chosen for effective treatment of different types of prostate cancer. Therefore, there is an urgent need to understand the mechanisms for these distinct subgroups of prostate cancer and for augmenting existing classification methods such as Gleason grading and TNM staging.

Like other cancers, prostate cancer is characterized by frequent genomic copy number (CN) alterations even at early stages. Using cytogenetic, loss of heterozygosity (LOH), comparative genomic hybridization (CGH) and other approaches, deletions and gains in the genome of prostate cancers have been identified from a number of studies (Kibel et al., 2000; Dong, 2001, 2006; Chu, et al., 2003; Clark

Supported by: National Institutes of Health, Grants numbers: CA106523 (to J.X.) and CA95052 (to J.X.); Department of Defense, Grants numbers: PC051264 (to J.X.).

*Correspondence to: Dr. Jianfeng Xu, Medical Center Blvd, Winston-Salem, NC 27157, USA. E-mail: jxu@wfubmc.edu or Dr. William B. Isaacs, Marburg 115, 600 N. Wolfe Street, Baltimore, MD 21287, USA. E-mail: wisaacs@jhmi.edu.

Received 9 June 2006; Accepted 5 July 2006
DOI 10.1002/gcc.20369
Published online 8 August 2006 in Wiley InterScience (www.interscience.wiley.com).

© 2006 Wiley-Liss, Inc.

WILEY InterScience® DISCOVER SOMETHING GREAT

et al., 2003; Dumur et al., 2003; Lieberfarb et al., 2003; Paris et al., 2003, 2004, 2005; van Dekken et al., 2003, 2004; Strohmeyer et al., 2004; Teixeira et al., 2004; Watson et al., 2004; Wolf et al., 2004; Yano et al., 2004; Kasahara et al., 2005; van Duin et al., 2005; Postma et al., 2006; Saramaki et al., 2006). Although several candidate tumor suppressors and oncogenes have been identified, the vast majority of cancer-associated genes involved in these genomic CN alterations are yet to be identified. This gap is partially due to lack of high-resolution and high-throughput methods to effectively pinpoint genomic CN alterations at the gene level in a large number of prostate cancers.

Genome-wide analyses with high-resolution have recently been developed for assessing genomic CN alterations. Representational oligonucleotide microarray analysis (ROMA), with an average resolution of about 35 kb, has been used to identify both genomic abnormalities in tumor cells and CN polymorphisms (CNPs) in the normal genome (Lucito et al., 2003; Sebat et al., 2004). Affymetrix 100K single nucleotide polymorphism (SNP) mapping array, with an average resolution of about 24 kb, has recently been demonstrated to be very effective in association studies of SNPs and in the identification of genomic CN alterations including CNPs (Garraway et al., 2005; Klein et al., 2005; Slater et al., 2005; Zhao et al., 2005; Liu et al., in press). In the present study, we report deletions and gains in prostate cancer genomes detected in DNA samples isolated from 22 pairs of tumor and matched nonmalignant tissues using Affymetrix 100K SNP mapping arrays.

MATERIALS AND METHODS

Study Subjects

All subjects in this study were prostate cancer patients undergoing radical prostatectomy (RP) for treatment of clinically localized disease at the Johns Hopkins Hospital. We selected 22 subjects from whom genomic DNA of sufficient amount (>5 μg) and purity (>70% cancer cells for cancer specimens, no detectable cancer cells for normal samples) could be obtained by macrodissection of matched nonmalignant (hereafter referred to as "normal") and cancer containing areas of prostate tissue as determined by histological evaluation of H&E stained frozen sections of snap frozen RP specimens. Genomic DNA was isolated from trimmed frozen tissues as previously described (Bova et al., 1993). DNA samples from prostate cancers meeting the same purity criterion and prepared in an identical fashion from an additional 69 patients were included in this study to estimate the frequency of identified deletions and gains using qPCR.

Affymetrix 100K SNP Mapping Array

The Affymetrix 100K SNP mapping array includes 116,204 SNPs in two chips, Xba240 and Hind240. The chips and reagents were obtained from Affymetrix and the assays were carried out according to the manufacturer's instructions. Briefly, 250 ng of genomic DNA were digested with either *Hind*III or *Xba*I and then ligated to adapters that recognize the cohesive four base-pair (bp) overhangs. A generic primer that recognizes the adapter sequence was used to amplify adapter-ligated DNA fragments with PCR conditions optimized to preferentially amplify fragments in the 250–2,000 bp size range in a GeneAmp PCR System 9700 (Applied Biosystems, Foster City, CA). After purification with a Qiagen MinElute 96 UF PCR purification system, a total of 40 μg of PCR product was fragmented and a sample of about 2.9 μg was visualized on a 4% TBE agarose gel to confirm that the average size was smaller than 180 bp. The fragmented DNA was then labeled with biotin and hybridized to the GeneChip Mapping 100K Set for 17 hr. We washed and stained the arrays using the Affymetrix fluidics Station 450 and scanned the arrays using a GeneChip Scanner 3000 G7 (Affymetrix, Santa Clara, CA). The Affymetrix GeneChip® Operating Software (GCOS) collected and extracted feature data from Affymetrix GeneChip® Scanners. The GeneChip Genotyping analysis software (GTYPE) was used to analyze feature intensity data stored in the GCOS Database, and provided high-throughput and accurate genotyping analysis.

DNA CN and Classification of Deletions and Gains

DNA CNs were calculated based on allele intensity (the sum of both allele intensity) of each SNP probe on the 100K SNP mapping array using three different software packages; Chromosome CN Analysis Tool (CNAT, Huang et al., 2004; Slater et al., 2005) version 3.0, CN Analyzer for Affymetrix GeneChip (CNAG, Nannya et al., 2005), and dChip analyzer (dChip, Lin et al., 2004). Deletions and gains were defined based on DNA CNs of 100K SNPs using a set of working criteria, which is implemented in an in-house script.

Quantitative Real Time PCR (qPCR)

A subset of putative deletions and gains were subjected to confirmation by quantitative real-time PCR (qPCR) using the ABI Prism 7000 Sequence

TABLE 1. Characteristics of Study Subjects

Subject ID	Gleason sum	Gland weight (g)	Age at surgery	Race
6-675	6	49	57	European American
6-795	6	47	51	African American
6-800	6	37	61	European American
6-816	6	42	62	European American
6-978	6	46	62	European American
6-1019	6	50	53	European American
6-1048	6	53	44	European American
6-1203	6	49	60	European American
7-372	7	93	69	European American
7-700	7	49	66	European American
7-721	7	42	58	European American
7-782	7	42	58	European American
7-814	7	57	67	European American
7-938	7	74	53	European American
7-989	7	64	58	European American
7-994	7	58	53	African American
8-535	8	55	55	European American
8-541	8	50	67	European American
8-780	8	60	53	European American
8-1070	8	55	66	European American
9-401	9	40	NA	European American
9-731	9	45	52	European American

Detection System. Primers were designed using Primer Express 1.5 software from Applied Biosystems. Amplicons were designed against the putatively altered locus and a control locus (GAPDH) with DNA CN of 2. The PCR kinetics at the control locus was used for controlling sample-to-sample differences in genomic DNA purity and concentration. Three concentrations of each genomic DNA sample (20, 10, and 5 ng) were assayed in duplicate, using each pair of real-time PCR primers. PCRs were prepared as follows: in 20 μl, we combined 2 μl of genomic DNA, 0.05 μM of each primer, and SYBR-Green PCR Master Mix from Applied Biosystems. PCRs were performed as follows: 95°C for 10 min, followed by 40 cycles at 95°C for 20 sec, and 60°C for 1 min. An additional cycle of 95°C for 15 sec, 60°C for 20 sec, and 95°C for 15 sec was run at the end to measure the dissociation curve for quality control. We used the Sequence Detection Software (SDS) for PCR baseline subtraction and exported the threshold cycle number (Ct) data for analysis. Ct values of the control (x-axis) and test (y-axis) amplicons for the three dilutions of each DNA sample were plotted. The differences in the Ct values (ΔCt) between tumor and matched normal DNA are used to infer DNA CN. ΔCt is approximately equal to $-\log 2(f)$ assuming PCR efficiency is 100%, where f is the ratio of DNA amount between tumor and normal samples. ΔCt is equal to one if f is 0.5 for a hemizygous deletion and ΔCt is equal to −0.58 if f is 1.5 for a three copies of DNA. Because of contamination of normal DNA in macrodissected tumor DNA in our study, the exact ΔCt values for deletions and gains are uncertain. Assuming 25–40% normal DNA contamination, the ΔCt for a hemizygous deletion is between 0.68 and 0.51, and for three copies of DNA is between −0.46 and −0.38.

Statistical Analysis

Kruskal-Wallis test was performed to assess difference in the numbers of DNA CN alterations among the four Gleason groups. The difference in the proportions of SNPs that are located within genes for SNPs involved in the altered regions and for all of the SNPs in the 100K SNP array was tested by the two sample proportion test (two-sided).

RESULTS

Detection of DNA CN Alterations in Macrodissected Prostate Tumors

Characteristics of the 22 subjects, including race, age at surgery, Gleason sum, and pathologic stage are presented in Table 1. DNA samples prepared in a similar fashion from prostate cancers of an additional 69 patients were used in qPCR analysis to estimate frequencies of alterations found in SNP based studies.

We obtained excellent SNP call rates using the 100K SNP mapping array for DNA samples isolated from tumor tissues (average of 97.34%) and

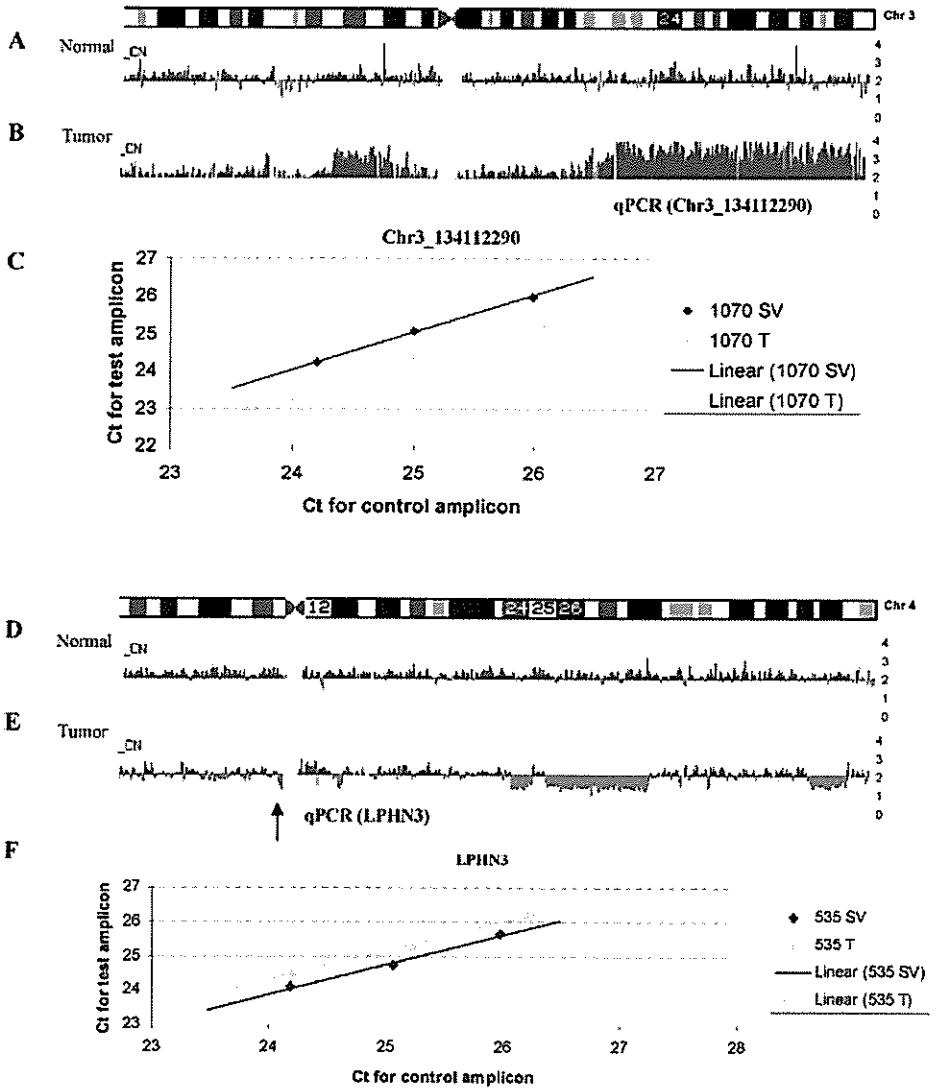

Figure 1. Examples of genomic gains on chromosome 3 and deletions on chromosome 4 identified in tumor genomes using Affymetrix100K SNP mapping array. While DNA CNs from matched normal tissues were ~2 on chromosome 3 (Fig 1A) and on chromosome 4 (Fig 1D), DNA CNs were increased (CN = 3–4) in multiple regions on chromosome 3 (Fig. 1B) and decreased (CN = 1) in multiple regions on chromosome 4 (Fig. 1E). Results from qPCR analyses were consistent with that of SNP mapping array. Compared with the DNA CN of normal DNA, DNA CN in tumor DNA was higher on chromosome 3 (Fig. 1C) and lower on chromosome 4 (Fig. 1F). [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

from matched normal tissues (average of 97.45%). The high SNP call rates suggest high quality allele intensity data. DNA CN at each SNP was estimated from the allele intensity data using three different software packages, CNAT, CNAG, and dChip. Similar results for DNA CNs were obtained from these analyses; however, we primarily described the results obtained from CNAT.

DNA CN alterations can be detected in DNA from macrodissected cancer samples using the 100K SNP mapping array. Figure 1 presents typical results of DNA CNs estimated from tumor DNA and from the matched normal DNA. While DNA CNs from matched normal tissues were ~2 for this subject on chromosome 3 (Fig. 1A) and on chromosomes 4 (Fig. 1D), cancer DNA CNs were increased (CN = 3–4) in multiple regions on chromosome 3 (Fig. 1B) and decreased (CN = 1) in multiple regions on chromosome 4 (Fig. 1E). We chose one region each from chromosome 3 and 4 and performed qPCR. Results from the qPCR analyses were consistent with that of SNP mapping array. Compared with the DNA CN of normal DNA, DNA CN in cancer DNA was higher on chromosome 3 (Fig. 1C) and lower on

chromosome 4 (Fig. 1F). Note that this particular deleted region was relatively small (<1 Mb), but can be detected by both SNP mapping array and qPCR. These results demonstrated that Affymetrix 100K SNP mapping array can be used to detect CN alterations in DNA prepared from macrodissected prostate tumors.

Working Criteria for Defining Genome-Wide Deletions and Gains

There are several factors that are known to prevent accurate estimates of DNA CNs. First, recent studies comparing specific loci among different individuals have reported substantial CN variation in germline DNA (Iafrate et al., 2004; Sebat et al., 2004; Sharp et al., 2005; Slater et al., 2005; Tuzun et al., 2005; Conrad et al., 2006; Hinds et al., 2006; Liu et al., in press; McCarroll et al., 2006). Second, random noise in allele intensity at each SNP may yield unreliable information on DNA alterations at the region. Third, polymorphisms/mutations in sequences that interfere with restriction enzyme digestion or probe hybridization may result, respectively, in longer DNA fragments which are likely to reduce the PCR yield of this region or limit hybridization rates. Ultimately these may lead to altered allele intensity and incorrect inclusion or exclusion of CN alterations.

With these issues in mind, we then set out to establish overall approaches to accurately infer putative deletions and gains in the whole genome based on the DNA CNs of the 116,204 SNP probes in the 100K mapping array. To minimize the impact of individual variability in germline DNA CN while improving our detection of somatic (versus germline) alterations in the cancer genome, we performed simultaneous SNP analyses of tumor DNA and matched normal tissue DNA, and then used the ratio of DNA CNs between tumor and normal samples as the primary variable. To limit the potential for artifactual deletions due to additional point mutations at restriction enzyme sites that result in fragments that are too long for consistent PCR amplification, we set a minimum physical length of 2 kb for putative CN alterations. To reduce random noise in allele intensity at individual SNPs, we estimated DNA CN based on multiple flanking SNPs in the region using the default 500-kb setting of CNAT software to perform *g*enome *s*mooth *a*verage *c*opy *n*umber (GSACN). However, we also used *s*ingle *p*oint *c*opy *n*umber (SPCN) to assist GSACN in defining alterations, to search for unique, high resolution information regarding DNA CN at each specific genomic position. Collectively, these approaches minimize factors known to influence the estimates of DNA CNs, while improving the accuracy of inferring DNA alterations.

Drawing from our overall approaches, we then experimented with a set of initial working criteria to define putative deletions and gains. To do this, we compared the results of identified alterations from these different criteria based on which ones: (1) gave the highest proportion of recurrent alterations ($\geq$2 subjects) among all of the identified alterations, and (2) could be 100% confirmed by quantitative real-time PCR analyses in eight candidate regions of deletion and four regions of gain. Based on these comparisons, we then selected the working criteria, one each for deletions and gains, for use as the primary analyses in this study. For deletions, the working criteria are a minimum four consecutive SNPs with at least three of them having the following characteristics: the GSACN and SPCN ratios of tumor/matched normal <0.75; the GSACN of the tumor DNA <1.9 for autosomal chromosomes, or <0.9 for X chromosome, and the minimum physical length of the putative deletion $\geq$2 kb. For gains, the working criteria are a minimum four consecutive SNPs with at least three of them having the following characteristics: the GSACN and SPCN ratios of tumor/match normal >1.4; the GSACN of the tumor DNA >2.7 for autosomal chromosomes, or >1.7 for X chromosome, and the minimum physical length of the putative gains $\geq$2 kb.

Comprehensive Assessment of Deletions and Gains in Prostate Tumors

We applied these two working criteria to examine the DNA CN data of 100K SNPs among 22 paired tumor/normal DNA samples. As shown in Supplementary Figure 1, we observed several large-scale CN alterations that are consistent with previous findings from cytogenetic and CGH studies, including deletions at 5q, 6q, 8p, 10q, 12p, 13q, and 16q and gains at 3, 7, and 8q. In addition, we were able to detect a number of smaller scale CN alterations. We found 863 putative deletions and 495 putative gains in these tumor genomes, of which 355 deletions (41%) and 223 gains (45%) are recurrent. The chromosomal band, the starting and ending positions, the number of SNP probes involved, the number of tumors involved, and the known genes within the regions of all of the recurrent alterations are presented in Supplementary Table 1a and 1b. The altered regions where at least four tumors were involved are presented in Table 2. The most fre-

quently deleted regions were 10q23.3 and 13q21.31, where seven tumors showed reductions in DNA CN compared to their normal counterparts, consistent with both hemizygous and homozygous deletion (see later). The other commonly deleted regions were 3q26.33, 4q28.3, and 12q21.32 among six tumors, and 6q14.3, 8p22 (at ~14 Mb), 8p22 (at ~16 Mb), 8p11.1, 8p11.2, 16q22.1, 13q31.1, and 16p11.2, among five tumors. The most frequently observed gains were detected on 11q13.5, where seven tumors were involved. The other regions with common gains were 7q22.1 and 16q22.1, where six tumors were involved, and 1p36, 7p22.2, and 22q13.31, where five tumors were involved.

Many of these frequent alterations have been previously reported by various methods. In fact, all 20 frequent deletions (>10%) on autosomal chromosomes described in the review by Dong (2001) were observed in our study and were found in at least two of our samples. In addition, we identified many regions that have not been previously reported as frequent changes. For example, among the 43 regions where at least four of the 22 tumors (>18%) were found in our study to contain a deletion (Table 2), seven of these regions were novel, including 3q26.33, 4q32.2, 4q34.1, 5q12.2, 6q24.3, 9q31.1, and 13q31.1. As expected, the sizes of these seven novel deleted regions were smaller (median size of 206,587 bp) than that of 37 previously known deleted regions (median size of 392,609 bp). This difference in size was statistically significant, $P = 0.01$ (nonparametric rank test), thus demonstrating the advantage of this high-resolution method in detecting smaller alterations.

Importantly, we found significant differences in the number of deletions and gains among tumors of different Gleason scores (Fig. 2, Table 3). For tumors with Gleason 6–8, we observed a trend of more DNA CN alterations in the tumors with higher Gleason sums. For example, the median numbers of deletions were 17.5, 50, and 205 for the tumors of Gleason 6, 7, and 8, respectively. The median numbers of gains were 1.5, 35, and 94 for the tumors of Gleason 6, 7, and 8, respectively. It is interesting to note that somatic CN alterations are less common in the two Gleason 9 tumors in our study. An average of 25.5 deletions and no gains were found in these two subjects. While these observations were based on only two subjects, this trend appeared to hold in our analysis of a larger number of Gleason 9 samples at the *PTEN* locus (see the section below "PTEN hemizygous and homozygous deletion in primary prostate tumors"). An examination of the tumor histology for the specimens from which these two DNA samples were isolated indicates that this lower frequency of CN alterations was not due to relatively lower tumor purity of these samples (data not shown). Further studies in a larger number of samples are needed to obtain a better estimate of genomic CN alterations in the tumors of Gleason 9 or higher.

Genes Implicated in the Regions of DNA CN Alterations

While multiple recurrent DNA CN alterations were located between genes, more than half of them (58%) involved genes, either completely or in part (Table 4). Specifically, the vast majority (74%) of the regions with recurrent CN gains involved genes; 571 known genes were located in the 223 recurrent gained regions. In contrast, only about one half (49%) of the recurrent deleted regions involved genes; 459 known genes were located in the 355 recurrent deleted regions. The difference in the proportion of alterations involving genes between gain and loss events was statistically significant ($Z = 6.00$, $P < 0.000001$).

To further test whether the DNA CN alterations preferentially target genes, we compared the proportion of the altered SNP probes that are located within genes with that of all of the SNP probes on the 100K array. Among the 116,204 SNPs, 41,959, corresponding to 36%, are located within genes, as defined by the termini of the 5′ and 3′UTRs. In comparison, both recurrent gains (74%, $Z = 11.76$, $P < 0.00001$) and deletions (49%, $Z = 4.94$, $P < 0.00001$) were significantly more likely to occur within genes than the average in the genome (36%). When the proportions of SNPs that are located within genes are estimated separately for alterations as a function of their frequency of occurrence in multiple tumors, we found that the proportions were higher for SNPs involved in alterations that were observed ≥4 subjects, significantly higher than the average in the genome (36%); 45% for the deletions ($Z = 2.65$, $P = 0.008$), and 47% for the gains ($Z = 2.90$, $P = 0.004$). These results indirectly suggest that recurrent CN alterations, both gain and loss, preferentially target gene-containing intervals, consistent with a role for selection of cells with increasing numbers of specific gene dosage alterations as prostate cancers initiate and progress.

PTEN Hemizygous and Homozygous Deletion in Primary Prostate Tumors

The high-resolution SNP array provides an excellent tool to better define altered regions

TABLE 2. Putative Deleted and Amplified Regions Identified in 22 Primary Tumors Using 100K SNP Mapping Panel

Chromosomal band	Implicated region (bp)			No. of implicated		Known genes in the regions (UCSC)
	Start	End	Size	SNPs	SNPs	
Deleted regions						
10q23.2-23.31	89,308,314	90,653,819	1,345,505	84	7	PAPSS2, PTEN, ATAD1, C10orf59, LIPF, ANKRD22
13q21.31	60,860,097	62,682,287	1,822,190	81	7	PCDH20
3q26.33	182,126,289	182,248,076	121,787	5	6	FXR1, DNAJC19
4q28.3	136,423,989	136,677,252	253,263	16	6	—
12q21.32	86,008,606	86,199,128	190,522	10	6	—
8p11.1	43,218,212	43,312,864	94,652	6	5	POTE8
6q14.3	86,797,435	87,171,740	374,305	14	5	—
8p22	14,201,821	14,903,851	702,030	77	5	SGCZ, TUSC3
8p22	16,181,469	16,758,362	576,893	38	5	—
8p12-11.21	37,608,457	40,420,045	2,811,588	60	5	ZNF703, SPFH2, PROSAC, GPR124, BRF2, RAB11FIP1, ADRB3, EIF4EBP1, ASH2L, STAR, LSM1, BAG4, DDHD2, PPAPDC1B, WHSC1L1, LETM2, AF173898, FGFR1, FLJ43582, TACC1, HTRA4, TM2D2, ADAM9, ADAM32, AK129810, BC026083, BC047448, BC067864, AK128178, ADAM18, ADAM2, INDO, INDOL1, C8orf4
13q31.1	80,591,525	80,711,441	119,916	10	5	—
16q22.1	67,987,772	68,461,639	473,867	8	5	CYB5B, NFAT5, NQO1, NOB1P, WWP2
16q22.3-23.1	72,745,821	73,507,260	761,439	19	5	PSMD7, AK124154, LOC497190, MGC34761, LOC348174, GLG1, AK131501, RFWD3, FA2H, WDR59
2q22.2	143,455,911	143,574,902	118,991	5	4	KYNU
17q21.31	39,876,004	40,523,389	647,385	5	4	KIAA0553, FZD2, CCDC43, DBF4B, AY358101, ADAM11, GJA7, HIGD1B, EFTUD2, LOC388389, GFAP, AK124465, C1QL1, DCAKD, NMT1
4q34.1	176,402,528	176,589,178	186,650	6	4	—
5q12.2	63,386,968	63,692,865	305,897	5	4	RNF180, BC101279
3p12.2	83,456,290	83,694,096	237,806	7	4	—
4q32.2	163,606,393	163,880,053	273,660	16	4	—
5q21.3	104,253,771	105,557,038	1,303,267	29	4	—
5q23.1	119,189,176	119,485,637	296,461	10	4	—
6q14.1	80,611,908	80,870,799	258,891	18	4	ELOVL4, TTK
6q16.1	95,413,900	96,043,261	629,361	15	4	—
6q21	110,047,191	110,226,824	179,633	14	4	C6orf199, KIAA0274
6q24.3	146,164,446	146,394,383	229,937	6	4	FBXO30, SHPRH, GRM1
8p23.2	4,852,294	4,993,234	140,940	12	4	—
8p22	15,060,320	15,328,979	268,659	14	4	—
8p22	15,378,001	15,700,057	322,056	25	4	TUSC3
8p22	17,855,816	18,169,577	313,761	23	4	PCM1, ASAH1, NAT1
8p21.2	27,633,311	28,212,897	579,586	30	4	FLJ10853, SCARA3, MGC45780, PBK, ELP3, PNOC
8p12	32,318,980	32,632,781	313,801	32	4	NRG1
8p12	35,009,528	35,271,112	261,584	14	4	—
9q31.1	102,478,986	102,685,573	206,587	9	4	—
13q14.12	45,310,109	45,731,718	421,609	26	4	KIAA0853, CPB2, LCP1, LOC220416
13q21.1	52,895,341	53,484,638	589,297	36	4	—

(Continued)

TABLE 2. Putative Deleted and Amplified Regions Identified in 22 Primary Tumors Using 100K SNP Mapping Panel (Continued)

Chromosomal band	Implicated region (bp)			No. of implicated		Known genes in the regions (UCSC)
	Start	End	Size	SNPs	SNPs	
13q21.1	54,585,799	54,780,048	194,249	15	4	—
13q21.1	56,263,149	57,516,869	1,253,720	42	4	FLJ40296, PCDH17
13q21.32	65,751,760	66,201,555	449,795	36	4	PCDH9
13q21.33	70,000,334	70,411,247	410,913	24	4	—
13q21.33	70,639,393	71,364,849	725,456	45	4	DACH1
13q21.33	71,438,158	71,624,921	186,763	14	4	—
13q22.1	72,183,522	72,836,342	652,820	28	4	AK025522, FLJ22624, AK095410, KIAA1008, C13orf24, KLF5
16q23.1	74,948,793	75,095,527	146,734	18	4	CNTNAP4
Amplified regions						
11q13.5	75,434,322	75,922,141	487,819	6	7	UVRAG, WNT11, PRKRIR, BC040665, LOC387790, EMFY
7q22.1	101,238,531	101,259,615	21,084	4	6	CULT1
16q22.1	66,582,351	67,026,799	444,448	7	6	DPEP2, CR625664, DOX28,DUS2L, NFATC3, RBM35B, LYPLA3, SLC7A6, SLC7A6OS, PRMT7, AK123945, SMPD3, AK128261
1p36	2,960,027	3,326,028	366,001	12	5	PRDM16
7p22.2	2,545,672	2,779,870	234,198	12	5	C7Orf27, IQCE, TTYH3, AM21, GAN12
22q13.31	42,873,845	43,132,573	258,728	6	5	PARVB, PARVG, BC104183
7q36.3	154,307,728	154,780,169	472,441	6	4	DPP6, PAXIP1, HTR5A, INSIG1
8q13.3	71,028,222	71,500,739	472,517	22	4	FRDM14, NCOA2
1p36.31	5941712	6666834	725122	7	4	NPNP4, KCNAB2, CHD5, AK094219, RPL22, FLJ46380, BC034459, C1orf188, AK128450, ICMT, MGC40168, HES3, GPR153, ACOT7, AY358179, HES2, ESPN, TNFRSF25, PLEKHG5, NOL9, TAS1R1, HKR3, KLHL21, PHF13, THAP3, DNAJC11
1q42.3	232938058	233126758	188700	17	4	AF193050
3q22.3	139573887	140078942	505055	11	4	MARS, FAM62C, CEP70, FAIM, PIK3CB
3q26.2	170988843	171659151	670308	12	4	LRRC34, LRRC31, SAMD7, TLOC1, GPR160, PHC3, AK095225, PRKCI, SKIL, CLDN11
3q26.32	177666873	178105323	438450	13	4	—
5p15.33	3101932	3189229	87297	7	4	—
5q35.3	180226111	180607628	381517	8	4	BTNL8, BTNL3, BTNL9, TRIM7, TRIM41, AX775803, AX775797, AX775789, AX775791, GNB2L1
7p15.2	25276343	25599789	323446	29	4	—
7p12.3	47225158	47832160	607002	13	4	TNS3, AK126096, BC007354, PKD1L1, FL21075
7p11.2	56479844	57174263	694419	7	4	LOC401357
7q11.21	62940814	63362316	421502	11	4	ZNF679
7q11.23	76478717	76636664	157947	10	4	LOC389571, KIAA1505
7q21.12	86647717	87296218	648501	22	4	DMTF1, C7orf23, TP53AP1, CROT, ABCB4, ABCB1, RPIB9
7q36.3	157034271	158246990	1212719	8	4	PTPRN2, AK126705, AK057320, AK127222, LUZP5, FAM62B
8q11.23	53174241	53532128	357887	21	4	ST18
8q12.3	64607420	64947590	340170	11	4	—
8q13.2	69108760	69369253	260493	21	4	DEPDC2

(Continued)

TABLE 2. Putative Deleted and Amplified Regions Identified in 22 Primary Tumors Using 100K SNP Mapping Panel (Continued)

Chromosomal band	Implicated region (bp)			No. of implicated		Known genes in the regions (UCSC)
	Start	End	Size	SNPs	SNPs	
8q21.11	75495378	75880359	384981	12	4	—
8q22.1	96918711	97386585	467874	14	4	GDF6, UQCRB, MTERFD1, PTDSS1
8q24.21	127511831	128258556	746725	48	4	FAM84B
8q24.21	129018405	129501894	483489	34	4	TMEM75
8q24.22	134424597	135082614	658017	31	4	ST3GL1, AK002210
8q24.22-24.23	136301294	136599508	298214	8	4	KHDRBS3
10p15.3	284953	737115	452162	9	4	ZMYND11,DQ335455, DIP2C, AK130224, C10orf106, AK096013
20q13.31	54998112	55225345	227233	7	4	BMP7

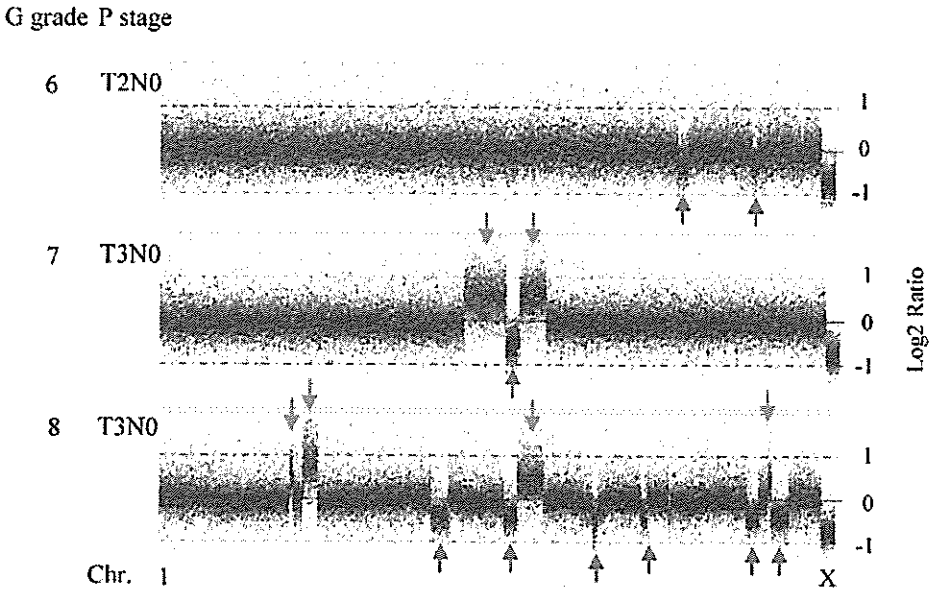

Figure 2. Whole-genome comparison of genomic aberrations among prostate tumors with different Gleason scores at different pathological stages. Chr, chromosome. Green arrow, deletion. Red arrow, amplification. Only graphically obvious aberrations are marked. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

shared by multiple tumors, thereby improving the ability to identify specific genes that may be driving the selection of these alterations. We examined the genes implicated at 10q23, the most commonly deleted region in the current study population. Although the entire affected region spanned ~1.3 Mb, only two in-gene SNP probes, spanning 10,053 bp, were deleted in each of these seven tumors (Figs. 3A–C). *PTEN* was the only gene residing within these two SNPs.

To confirm the deletion status at *PTEN*, we performed a qPCR analysis for these 22 pairs of tumor and normal DNA samples using a probe at position 89,675,748 bp, between the two implicated SNPs. The results from the qPCR analysis were consistent with that from Affymetrix SNP mapping method. The differences in the Ct values (ΔCt) between tumor and matched normal among these subjects were >0.51, suggesting a *PTEN* deletion in these tumors. Among the 7 *PTEN* deleted tumors, four were Gleason 8, two were Gleason 7, and one was Gleason 6 (Fig. 3).

To better estimate the frequency of the *PTEN* deletion in primary prostate tumors, we performed qPCR analysis in another population of 69 tumors that had not been analyzed by the 100K SNP array. Among a total of 91 primary prostate tumors examined, 32 tumors (32/91 = 35%) had the deletion,

TABLE 3. DNA Copy Number Alterations Among Tumors with Different Gleason Sum

Gleason sum	No. of subjects	Median copy number alterations			Median recurrent copy number alterations		
		Deletions	Gains	Both	Deletions	Gains	Both
6	8	17.5	1.5	26.5	5.5	0.5	7
7	8	50	35	74	25	15	44
8	4	205	94	299	76.5	41.5	118
9	2	25.5	0	25.5	2	0	2
P-value[a]		0.04	0.006	0.02	0.005	0.01	0.01

[a]Kruskal–Wallis test was performed to test if there was a statistically significant difference of number of genetic alterations among the four Gleason score groups.

TABLE 4. Proportions of Recurrent Alterations Involving Genes

Type of alterations	No. of alterations	No. (%) of alterations involving genes	No. of genes involved in alterations
Deletions observed in			
≥4 subjects	44	25 (56.82%)	121
3 subjects	68	36 (52.94%)	110
2 subjects	243	112 (46.09%)	228
Sub total	355	173 (48.73%)	459
Gains observed in			
≥4 subjects	33	28 (84.84%)	127
3 subjects	48	42 (87.50%)	129
2 subjects	142	95 (66.90%)	315
Sub total	223	165 (73.99%)	571
Total	578	338 (58.47%)	1,030

defined as $\Delta Ct \geq 0.51$ (Table 5). Importantly, we continued to observe higher percentages of *PTEN* deletion in the tumors of Gleason 8 (7/10 = 70%) and Gleason 7 (21/51 = 41%), compared with the tumors of Gleason 6 (2/18 = 11%). Interestingly, consistent with the overall pattern of fewer somatic CN deletions and gains in two Gleason 9 tumors observed in our study, the *PTEN* deletion was relatively uncommon in the 12 Gleason 9 tumors, as only two subjects had a deletion (2/12 = 17%).

We attempted to further differentiate homozygous deletions from hemizygous deletions based on the ΔCt of the qPCR analysis between tested tumor and normal DNA samples. Because of various levels of normal DNA contamination in these macrodissected primary tumor DNA samples, the classification of homozygous and hemizygous deletions based on ΔCt was subject to errors. However, assuming a minimum of 25% normal DNA contamination in our tumor DNA, the maximum ΔCt for a hemizygous deletion is 0.68. Therefore, tumors with $\Delta Ct > 0.68$ were consistent with a homozygous deletion. Among the 32 *PTEN* deleted tumors, 13 could be classified as homozygous deletions (Table 5, Fig. 4). Interestingly, all of the *PTEN* homozygous deletions were found in Gleason 8 and 7 tumors, including six Gleason 8 tumors (6/10 = 60%) and seven Gleason 7 tumors (7/51 = 14%). None of the eighteen Gleason 6 tumors were classified as homozygous deletions, and neither were any of the twelve Gleason 9 tumors.

Common Deletions Between *ERG* and *TMPRSS2*

The high-resolution SNP array also provides an excellent tool for discovering novel alterations. One novel region of frequent recurrent deletion, at 21q22.2, was of particular interest. Using 500K GSACN, we found that six tumors had apparent deletions between 38 Mb and 41.7 Mb, corresponding approximately to the interval between the genes, *ERG* and *TMPRSS2*, recently demonstrated to be involved in common gene fusion events in prostate cancer (Tomlins et al., 2005). The GSACN ratio of tumor/normal of these six subjects is presented in Figure 5. On the basis of the data from GSACN and SPCN analyses, the boundaries of the deletions at the telomeric side appear to be within a small region, at least among five of the six tumors, involving four SNPs and spanning 217,769 bp. While the telomeric boundary may reside within *TMPRSS2*, it is difficult to infer the exact breakpoint because no SNP on the 100K array resides within *TMPRSS2*. In contrast, the boundaries of the deletions on the centromeric side appear to vary among the six tumors, involving 99 SNPs that span 1,119,680 bp. The deletion boundaries in three tumors are consistent with breakpoints within the *ERG*.

DISCUSSION

DNA CN alterations may provide important clues in identifying tumor suppressor genes and oncogenes. The ability to identify causal genes in the altered DNA CN regions largely depends on the size of altered region, which is partially related to the resolution of methods in detecting the DNA CN alterations. In this study, we demonstrated that

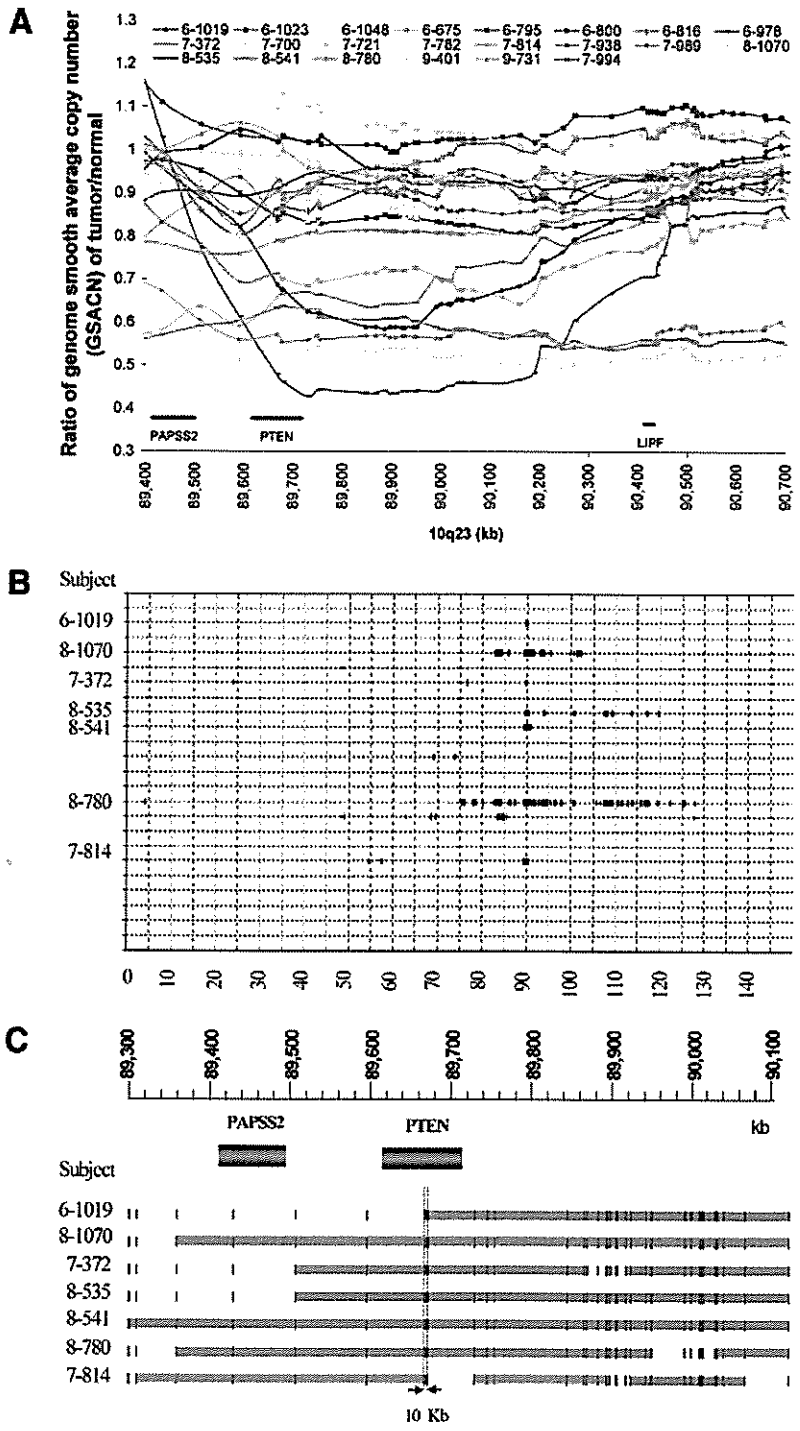

Figure 3. Analysis of DNA copy number alterations on Chromosome 10. (A) scatter plot of GSACN ratio of tumor/normal at each SNP locus from 89.4 to 90.7 Mb to display the alterations of DNA copy numbers at 10q23 among 22 prostate cancer subjects, each of which is labeled with Gleason score–tumor ID. Horizontal bars indicate the physical positions of PASS2, PTEN and LIPF. (B) putative deletions detected in 10 tumors. Each deletion is plotted as vertical bar at the specific physical location on chromosome 10. (C) overlapping analysis of recurrent deletions from 89.3 to 90.1 Mb on the physical map among seven subjects. Black vertical bar represents the location of each SNP. Deleted SNPs are high-lighted in red. Horizontal bars indicate the physical positions of PASS2 and PTEN. Dot vertical lines mark the maximum overlapping region deleted within PTEN with arrows indicating the size of the overlapping deletion. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

Affymetrix 100K SNP mapping array, with an average resolution of one SNP in 24 kb, provide a high resolution method to detect DNA CN alterations in the tumor genome. We obtained several important findings in our study by using this new method to examine 22 pairs of primary prostate cancers and matched normal tissues. We detected 355 recurrent deletions and 223 recurrent gains in these tumor genomes, many of which were novel, particularly those of smaller sizes alterations. We found significantly higher numbers of DNA CN alterations in tumors with higher Gleason scores

TABLE 5. Putative Homozygous and Heterozygous PTEN Deletions

Gleason sum	No. of subjects	No. (%) of PTEN deletion status		
		Normal	Hemizygous	Homozygous
6	18	16 (88.89)	2 (11.11)	0 (0)
7	51	30 (58.82)	14 (27.45)	7 (13.73)
8	10	3 (30.00)	1 (10.00)	6 (60.00)
9	12	10 (83.33)	2 (16.67)	0 (0)
Total	91	59 (64.83)	22 (24.18)	10 (10.99)

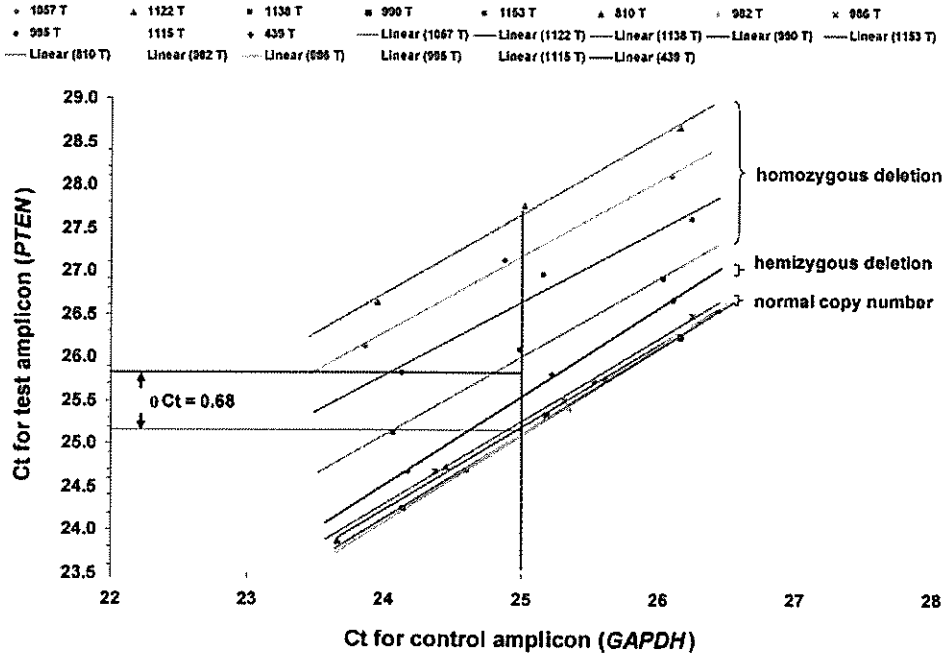

Figure 4. Validating PTEN deletion by real-time quantitative PCR. Ct values of the control (*x*-axis) and test (*y*-axis) amplicons for the three dilutions of each DNA sample were plotted against each other, and the offsets between best-fit lines for the samples along the test-amplicon axis at 25 Ct of the control-amplicon axis were measured. The offsets in the Ct values (ΔCt) between tumor DNAs with putative PTEN deletions and tumor DNAs with known normal PTEN are used to infer DNA copy number. A tumor DNA is defined as having PTEN hemizygous deletion when the ΔCt is less than 0.68 and as having PTEN homozygous deletion when the ΔCt is more than 0.68, assuming 25% normal DNA contamination. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

among tumors with Gleason sums 6, 7, and 8. We observed that most frequent DNA CN alterations of both types (deletion and gain) preferentially occurred within genes. Finally, we observed a novel and recurrent deletion between the *ERG* and *TMPRSS2* genes on chromosome 21, presumably related to the recently identified formation of fusion transcripts from these two genes. These findings demonstrated the advantages of this high-resolution method in detecting DNA CN and provided important clues for further classifying heterogeneous prostate tumors and identification of genes important in tumorigenesis.

Another important feature of our study is that we demonstrate the feasibility of SNP mapping arrays in detecting CN alterations in DNA isolated from macrodissected tumor tissues. While microdissection of tissue samples using techniques such as laser capture have an advantage of yielding potentially more homogeneous samples, with little contamination from nonneoplastic cells, this approach provides only limited amounts of DNA available for repeat and subsequent follow up analyses, and the results obtained reflect only the characteristics of the small number of cells analyzed. While being more susceptible to contamination by normal cells, the macrodissection method used in this study provides a larger quantity of DNA, and reflects the average genetic makeup of a much larger number of cells. Thus while CN alterations that may be specific for a small subset of cancer cells may not be detected, our detection of hun-

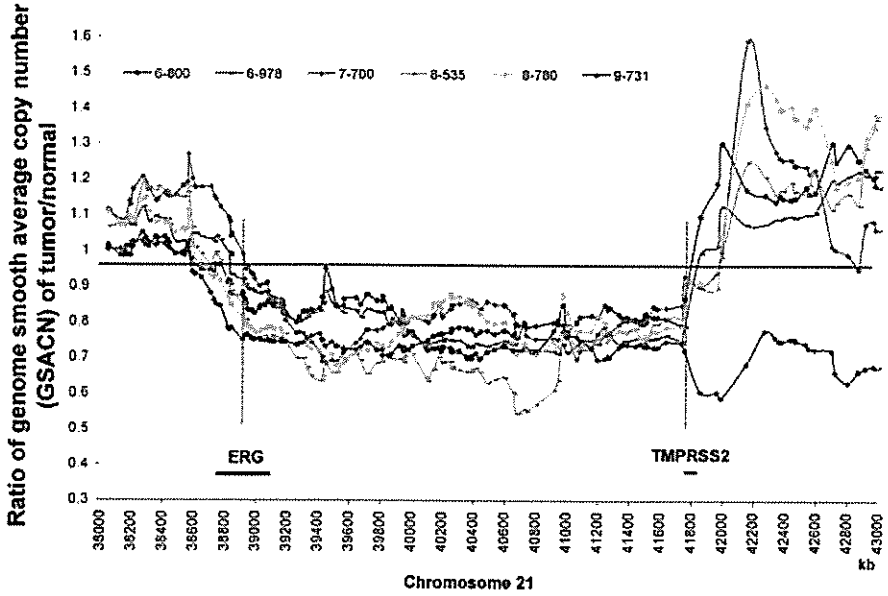

Figure 5. Analysis of DNA copy number alterations on Chromosome 21. The tumor/normal ratios of GSACN at each SNP locus from 38 to 43 Mb are plotted to illustrate the alterations of DNA copy numbers in six tumors. The tumor/normal ratios of GSACN of many SNPs in the regions between the two dot vertical lines indicate deletions between ERG and TMPRSS2 whose positions are marked as horizontal bars. [Color figure can be viewed in the online issue, which is available at www.interscience.wiley.com.]

dreds of gains and deletions, affecting both documented and novel loci, emphasizes the usefulness of this approach to efficiently identify and characterize recurrent CN changes present in the majority of tumor cells within a given prostate cancer lesion. To address the question of genetic heterogeneity within prostate cancers, additional studies using high resolution SNP arrays to characterize multiple DNA samples isolated from separate, small cell populations are needed and should be highly informative.

It is interesting to note that we did not use genotype information of the SNP arrays to perform LOH analysis. Theoretically, the ability to compare SNP genotypes of matched tumor and normal DNA is an advantage of our approach, and LOH information is critical in defining deletions and allelic imbalance alterations. However, in practice we found that Affymetrix SNP genotyping is extremely sensitive and is able to detect very small numbers of alleles; therefore this approach is not suitable for LOH analyses in studies such as ours, where there may be various amounts of normal DNA contamination from macrodissected tumor DNA.

Although we have demonstrated the utilities of 100K SNP mapping array in detecting DNA CN alterations in the genome, there remain many challenging issues regarding the accuracy of this method. One of the most important practical issues is the quality of allele intensity data of SNP probes, the basis for estimating DNA CN. The quality of allele intensity data, however, can be indirectly measured by SNP call rate. A low SNP call rate, for example <95%, is an indication of poor quality of allele intensity and/or contamination of other DNA sources, and therefore are not suitable for the analysis of DNA CN (data not shown). We therefore did not use any data from the arrays with a SNP call rate <95% in this study. Other issues that may affect the accuracy of inferring DNA CN in tumors include potential germline CNPs, mutations in the restriction enzyme sites, and random noise of allele intensity at some SNPs. We have taken several measures to minimize the effects of these issues. We used paired tumor and matched normal DNA and used their ratio of DNA CNs as the primary variable to lessen the effect of potential germline CNPs. To limit the potential for artifact deletions due to additional point mutations at restriction enzyme sites that result in fragments that are too long for consistent PCR amplification, we set a minimum physical length of 2 kb for putative CN alterations. We use both GSACN and SPCN in defining CN alterations to balance the unique information the random noise of allele intensity at each SNP. However, these cautious steps are not sufficient to ensure the accuracy of this method in inferring DNA CN. Independent

confirmations using other methods such as qPCR analysis are helpful to reduce false positive finding of DNA CN alterations.

The high resolution of the 100K SNP array facilitates the identification of smaller size of CN alterations. The fact that novel DNA CN alteration regions identified in this study are significantly smaller in size compared with the previously know altered regions demonstrated the advantage of high resolution. The fine resolution of SNP arrays also allows us to distinguish between different alterations within a small chromosomal region. For example, we detected several deletions in the regions of 13q21.31, 13q21.1, 13q21.32, and 13q21.33. These different deletions might be detected as a single deletion using lower-resolution methods. Therefore, this increased resolution should be considered when comparing the alteration frequencies obtained from this study versus those estimated from lower resolution methods. Furthermore, the high resolution of this method facilitates the identification of specific genes that may be driving the selection of these alterations and play roles in tumorigenesis. The fact that we can pinpoint two SNPs that are commonly implicated in all seven tumors in the 10q23 region and *PTEN* is the only genes residing within these two SNPs provides an excellent example.

The rate of *PTEN* deletion was observed to increase as the Gleason score of the tumors increase from 6 to 8, which is in consistent with the reports that *PTEN* loss and reduction of its expression was highly correlated with tumor of high Gleason score at more advance stages (McMenamin et al., 1999; Dreher et al., 2004; Majumder and Sellers et al., 2005). It is somewhat surprising that *PTEN* deletion was significantly lower in Gleason 9 tumors with only 17% in comparison to that in Gleason 7 and 8 tumors with 41 and 70% respectively (Table 5). In the study of Gleason score distribution of chromosomal aberrations in prostate cancer using CGH, Chu et al., (2003) found that the frequency of 10q25-qter deletion was also much lower in Gleason 6 and 9 tumors with 22.2 and 16.7%, respectively, in comparison to that in Gleason 7 and Gleason 8 tumors with 45.8 and 33.3% respectively.

The fusion of *TMPRSS2* and *ERG* at high frequency has been reported in prostate cancer (Tomlins et al., 2005; Soller et al., 2006). The region between *TMPRSS2* and *ERG* represents one of the novel frequently deleted regions in prostate tumors that we analyzed (Fig. 5, Supplementary Fig. 1, Chr21). It is interesting to note that we did not find evidence of the deletion between *TMPRSS2* and *ERG* in any of the 22 normal DNA samples, indicating that this deletion is tumor specific and somatic in origin. That the boundaries of the deletions on the centromeric side appear to vary among the different tumors is consistent with a recent finding of multiple transcripts with different sizes in different prostate adenocarcinoma samples (Soller et al., 2006). Although our results suggest that somatic deletions may be one of the mechanisms contributing to the commonly observed *ERG* and *TMPRSS2* gene fusions, further studies, using higher resolution SNP mapping arrays and other cytogenetic and molecular methods are needed to better define the boundaries of the deletions and to verify if and how the deletions result in the fusion of *ERG* and *TMPRSS2*.

In summary, we have identified both unique and recurrent CN alterations occurring across the genome of clinical prostate cancers using the 100K SNP array. The increased resolution and genome wide nature of these data provide a comprehensive and systematic approach to dissection of the alterations at the levels of specific genes and/or regulatory elements which characterize and/or may be driving the development of prostate cancer. This may in turn translate into more effective management of this heterogeneous disease.

REFERENCES

Bova GS, Fox WM, Epstein JI. 1993. Methods of radical prostatectomy specimen processing: A novel technique for harvesting fresh prostate cancer tissue and review of processing techniques. Mod Pathol 6:201–207.

Chu LW, Troncoso P, Johnston DA, Liang JC. 2003. Genetic markers useful for distinguishing between organ-confined and locally advanced prostate cancer. Genes Chromosomes Cancer 36:303–312.

Clark J, Edwards S, Feber A, Flohr P, John M, Giddings I, Crossland S, Stratton MR, Wooster R, Campbell C, Cooper CS. 2003. Genome-wide screening for complete genetic loss in prostate cancer by comparative hybridization onto c-DNA microarrays. Oncogene 22:1247–1252.

Conrad DF, Andrews TD, Carter NP, Hurles ME, Pritchard JK. 2006. A high-resolution survey of deletion polymorphism in the human genome. Nat Genet 38:75–81.

Cooperberg MR, Lubeck DP, Mehta SS, Carroll PR. CaPSURE. 2003. Time trends in clinical risk stratification for prostate cancer: Implications for outcomes (data from CaPSURE). J Urol 170:S21–S25; discussion S26, S27.

Dong JT. 2001. Chromosomal deletions and tumor suppressor genes in prostate cancer. Cancer Metastasis Rev 20:173–193.

Dong JT. 2006. Prevalent mutations in prostate cancer. J Cell Biochem 97:433–447.

Dreher T, Zentgraf H, Abel U, Kappeler A, Michel MS, Bleyl U, Grobholz R. 2004. Reduction of PTEN and p27kip1 expression correlates with tumor grade in prostate cancer. Analysis in radical prostatectomy specimens and needle biopsies. Virchows Arch 444:509–517.

Dumur CI, Dechsukhum C, Ware JL, Cofield SS, Best AM, Wilkinson DS, Garrett CT, Ferreira-Gonzalez A. 2003. Genome-wide detection of LOH in prostate cancer using human SNP microarray technology. Genomics 81:260–269.

Garraway LA, Widlund HR, Rubin MA, Getz G, Berger AJ, Ramaswamy S, Beroukhim R, Milner DA, Granter SR, Du J, Lee C, Wagner SN, Li C, Golub TR, Rimm DL, Meyerson ML, Fisher DE, Sellers WR. 2005. Integrative genomic analyses identify MITF as a lineage survival oncogene amplified in malignant melanoma. Nature 436:117–122.

Gonzalgo ML, Isaacs WB. 2003. Molecular pathways to prostate cancer. J Urol 170:2444–2452.
Hinds DA, Kloek AP, Jen M, Chen X, Frazer KA. 2006. Common deletions and SNPs are in linkage disequilibrium in the human genome. Nat Genet 38:82–85.
Huang J, Wei W, Zhang J, Liu G, Bignell GR, Stratton MR, Futreal PA, Wooster R, Jones KW, Shapero MH. 2004. Whole genome DNA copy number changes identified by high density oligonucleotide arrays. Hum Genomics 1:287–299.
Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C. 2004. Detection of large-scale variation in the human genome. Nat Genet 36:949–951.
Jemal A, Siegel R, Ward E, Murray T, Xu J, Smigal C, Thun MJ. 2006. Cancer statistics, 2006. CA Cancer J Clin 56:106–130.
Kasahara K, Taguchi T, Yamasaki I, Kamada M, Shuin T. 2005. Genetic changes in localized prostate cancer of Japanese patients shown by comparative genomic hybridization. Cancer Genet Cytogenet 159:84–88.
Kibel AS, Faith DA, Bova GS, Isaacs WB. 2000. Loss of heterozygosity at 12P12-13 in primary and metastatic prostate adenocarcinoma. J Urol 164:192–196.
Klein RJ, Zeiss C, Chew EY, Tsai JY, Sackler RS, Haynes C, Henning AK, SanGiovanni JP, Mane SM, Mayne ST, Bracken MB, Ferris FL, Ott J, Barnstable C, Hoh J. 2005. Complement factor H polymorphism in age-related macular degeneration. Science 308:385–389.
Lieberfarb ME, Lin M, Lechpammer M, Li C, Tanenbaum DM, Febbo PG, Wright RL, Shim J, Kantoff PW, Loda M, Meyerson M, Sellers WR. 2003. Genome-wide loss of heterozygosity analysis from laser capture microdissected prostate cancer using single nucleotide polymorphic allele (SNP) arrays and a novel bioinformatics platform dChipSNP. Cancer Res 63:4781–4785.
Lin M, Wei LJ, Sellers WR, Lieberfarb M, Wong WH, Li C. 2004. dChipSNP: Significance curve and clustering of SNP-array-based loss-of-heterozygosity data. Bioinformatics 20:1233–1240.
Liu W, Chang B, Li T, Dimitrov L, Kim S, Kim JW, Turner AR, Meyers DA, Trent JM, Zheng S, Isaacs WB, Xu J. Germline copy number polymorphisms involving larger than 100 kb are uncommon in normal subjects. Prostate (in press).
Lucito R, Healy J, Alexander J, Reiner A, Esposito D, Chi M, Rodgers L, Brady A, Sebat J, Troge J, West JA, Rostan S, Nguyen KC, Powers S, Ye KQ, Olshen A, Venkatraman E, Norton L, Wigler M. 2003. Representational oligonucleotide microarray analysis: A high-resolution method to detect genome copy number variation. Genome Res 13:2291–2305.
Majumder PK, Sellers WR. 2005. Akt-regulated pathways in prostate cancer. Oncogene 24:7465–7474.
McCarroll SA, Hadnott TN, Perry GH, Sabeti PC, Zody MC, Barrett JC, Dallaire S, Gabriel SB, Lee C, Daly MJ, Altshuler DM. International HapMap Consortium. 2006. Common deletion polymorphisms in the human genome. Nat Genet 38:86–92.
McMenamin ME, Soung P, Perera S, Kaplan I, Loda M, Sellers WR. 1999. Loss of PTEN expression in paraffin-embedded primary prostate cancer correlates with high Gleason score and advanced stage. Cancer Res 59:4291–4296.
Nannya Y, Sanada M, Nakazaki K, Hosoya N, Wang L, Hangaishi A, Kurokawa M, Chiba S, Bailey DK, Kennedy GC, Ogawa S. 2005. A robust algorithm for copy number detection using high-density oligonucleotide single nucleotide polymorphism genotyping arrays. Cancer Res 65:6071–6079.
Paris PL, Albertson DG, Alers JC, Andaya A, Carroll P, Fridly J, Jain AN, Kamkar S, Kowbel D, Krijtenburg PJ, Pinkel D, Schroder FH, Vissers KJ, Watson VJ, Wildhagen MF, Collins C, Van Dekken H. 2003. High-resolution analysis of paraffin-embedded and formalin-fixed prostate tumors using comparative genomic hybridization to genomic microarrays. Am J Pathol 62:763–770.
Paris PL, Andaya A, Fridlyand J, Jain AN, Weinberg V, Kowbel D, Brebner JH, Simko J, Watson JE, Volik S, Albertson DG, Pinkel D, Alers JC, van der Kwast TH, Vissers KJ, Schroder FH, Wildhagen MF, Febbo PG, Chinnaiyan AM, Pienta KJ, Carroll PR, Rubin MA, Collins C, van Dekken H. 2004. Whole genome scanning identifies genotypes associated with recurrence and metastasis in prostate tumors. Hum Mol Genet 13:1303–1313.
Paris PL, Weinberg V, Simko J, Andaya A, Albo G, Rubin MA, Carroll PR, Collins C. 2005. Preliminary evaluation of prostate cancer metastatic risk biomarkers. Int J Biol Markers 20:141–145.
Postma R, van Marion R, van Duin M, Vissers KJ, Wink JC, Schroder FH, Tanke HJ, Szuhai K, van der Kwast TH, van Dekken H. 2006. Array-based genomic analysis of screen-detected Gleason score 6 and 7 prostatic adenocarcinomas. Anticancer Res 26:1193–1200.
Saramaki OR, Porkka KP, Vessella RL, Visakorpi T. 2006. Genetic aberrations in prostate cancer by microarray analysis. Int J Cancer 119: 1322–1329.
Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, Chi M, Navin N, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A, Wigler M. 2004. Large-scale copy number polymorphism in the human genome. Science 305:525–528.
Sharp AJ, Locke DP, McGrath SD, Cheng Z, Bailey JA, Vallente RU, Pertz LM, Clark RA, Schwartz S, Segraves R, Oseroff VV, Albertson DG, Pinkel D, Eichler EE. 2005. Segmental duplications and copy-number variation in the human genome. Am J Hum Genet 77:78–88.
Slater HR, Bailey DK, Ren H, Cao M, Bell K, Nasioulas S, Henke R, Choo KH, Kennedy GC. 2005. High-resolution identification of chromosomal abnormalities using oligonucleotide arrays containing 116,204 SNPs. Am J Hum Genet 77:709–726.
Soller MJ, Isaksson M, Elfving P, Soller W, Lundgren R, Panagopoulos I. 2006. Confirmation of the high frequency of the *TMPRSS2/ERG* fusion gene in prostate cancer. Genes Chromosomes Cancer 45:717–719.
Strohmeyer DM, Berger AP, Moore DH II, Bartsch G, Klocker H, Carroll PR, Loening SA, Jensen RH. 2004. Genetic aberrations in prostate carcinoma detected by comparative genomic hybridization and microsatellite analysis: Association with progression and angiogenesis. Prostate 59:43–58.
Teixeira MR, Ribeiro FR, Eknaes M, Waehre H, Stenwig AE, Gierckskey KE, Heim S, Lothe RA. 2004. Genomic analysis of prostate carcinoma specimens obtained via ultrasound-guided needle biopsy may be of use in preoperative decision-making. Cancer 101:1786–1793.
Tomlins SA, Rhodes DR, Perner S, Dhanasekaran SM, Mehra R, Sun XW, Varambally S, Cao X, Tchinda J, Kuefer R, Lee C, Montie JE, Shah RB, Pienta KJ, Rubin MA, Chinnaiyan AM. 2005. Recurrent fusion of *TMPRSS2* and ETS transcription factor genes in prostate cancer. Science 310:644–648.
Tuzun E, Sharp AJ, Bailey JA, Kaul R, Morrison VA, Pertz LM, Haugen E, Hayden H, Albertson D, Pinkel D, Olson MV, Eichler EE. 2005. Fine-scale structural variation of the human genome. Nat Genet 37:727–732.
van Dekken H, Paris PL, Albertson DG, Alers JC, Andaya A, Kowbel D, van der Kwast TH, Pinkel D, Schroder FH, Vissers KJ, Wildhagen MF, Collins C. 2004. Evaluation of genetic patterns in different tumor areas of intermediate-grade prostatic adenocarcinomas by high-resolution genomic array analysis. Genes Chromosomes Cancer 39:249–256.
van Dekken H, Alers JC, Damen JA, Vissers KJ, Krijtenburg PJ, Hoedemaeker RF, Wildhagen MF, Hop WC, van der Kwast TH, Tanke HJ, Schroder FH. 2003. Genetic evaluation of localized prostate cancer in a cohort of forty patients: Gain of distal 8q discriminates between progressors and nonprogressors. Lab Invest 83:789–796.
van Duin M, van Marion R, Vissers K, Watson JE, van Weerden WM, Schroder FH, Hop WC, van der Kwast TH, Collins C, van Dekken H. 2005. High-resolution array comparative genomic hybridization of chromosome arm 8q: Evaluation of genetic progression markers for prostate cancer. Genes Chromosomes Cancer 44:438–449.
Watson JE, Doggett NA, Albertson DG, Andaya A, Chinnaiyan A, van Dekken H, Ginzinger D, Ha C, James K, Kamkar S, Kowbel D, Pinkel D, Schmitt L, Simko JP, Volik S, Weinberg VK, Paris PL, Collins C. 2004. Integration of high-resolution array comparative genomic hybridization analysis of chromosome 16q with expression array data refines common regions of loss at 16q23-qter and identifies underlying candidate tumor suppressor genes in prostate cancer. Oncogene 23:3487–3494.
Wolf M, Mousses S, Hautaniemi S, Karhu R, Huusko P, Allinen M, Elkahloun A, Monni O, Chen Y, Kallioniemi A, Kallioniemi OP. 2004. High-resolution analysis of gene copy number alterations in human prostate cancer using CGH on cDNA microarrays: Impact of copy number on gene expression. Neoplasia 6:240–247.
Yano S, Matsuyama H, Matsuda K, Matsumoto H, Yoshihiro S, Naito K. 2004. Accuracy of an array comparative genomic hybridization (CGH) technique in detecting DNA CN aberrations: Comparison with conventional CGH and loss of heterozygosity analysis in prostate cancer. Cancer Genet Cytogenet 150:122–127.
Zhao X, Weir BA, LaFramboise T, Lin M, Beroukhim R, Garraway L, Beheshti J, Lee JC, Naoki K, Richards WG, Sugarbaker D, Chen F, Rubin MA, Janne PA, Girard L, Minna J, Christiani D, Li C, Sellers WR, Meyerson M. 2005. Homozygous deletions and chromosome amplifications in human lung carcinomas revealed by single nucleotide polymorphism array analysis. Cancer Res 65:5561–5570.